5-2019

NMR Studies of Translational and Rotational Diffusion of a Phase Separating Eye Lens Protein: Bovine γB-crystallin

Aaron Thomas Fadden
atf9181@rit.edu

NMR Studies of Translational and Rotational Diffusion of a Phase Separating Eye Lens Protein: Bovine γB-crystallin

Aaron Thomas Fadden

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Chemistry in the
School of Chemistry and Materials Science,
College of Science
Rochester Institute of Technology
May 2019

SCHOOL OF CHEMISTRY AND MATERIALS SCIENCE
COLLEGE OF SCIENCE

ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

School of Chemistry and Materials Science

CERTIFICATE OF APPROVAL

M.S. DEGREE THESIS

This is to certify that the Master's Thesis of Aaron Thomas Fadden
has been examined and approved by the thesis
committee as satisfactory for the thesis required for
the M.S. degree in Chemistry at the convocation of May 2019.

_____
Dr. Lea Vacca Michel, *Thesis Co-Advisor*

_____
Dr. George Thurston, *Thesis Co-Advisor*

_____
Dr. Michael Coleman, *Graduate Program Director*

_____
Dr. Jeffrey Mills, *Committee Member*

_____
Dr. David Ross, *Committee Member*

I) ABSTRACT:

Cataracts are one of the leading causes of blindness in the world. Cataracts can occur when proteins (known as "crystallins") in the cytoplasm of the eye lens phase separate or associate, creating local fluctuations in the refractive index of the lens. Bovine γB-crystallin is analogous (in sequence, structure, and function) to the human γD-crystallin and can be isotopically labeled during growth with ^{13}C and ^{15}N when expressed in *Escherichia coli*. In this work, the two most important Brownian motions, rotational and translational diffusion, were measured under various temperatures and concentrations in an effort to better understand the intermolecular interactions and behavior of γB-crystallins in solution. Rotational Correlation Times (τ_c) were estimated from T_1/T_2 nuclear magnetic resonance (NMR) data, and pulsed field gradient NMR was used to measure translational diffusion. Preliminary results suggest that bovine γB-crystallins associate more with increased protein concentration and/or decreased temperatures. Both single and double exponential decays were used to fit the T_1 data, and the corresponding τ_c values were compared. Additionally, the average hydrodynamic radii of the molecules were approximated using the calculated τ_c values, which supports our hypothesis that the proteins are indeed associating. Diffusion coefficients were also measured at various concentrations, with preliminary results indicating that as concentration increases, diffusion coefficients decrease, supporting the theory that even small increases in protein concentration result in association of the γB-crystallins.

II) ACKNOWLEDGEMENTS

First of all, I would like to thank the RIT School of Chemistry and Materials Science, and specifically the Chemistry Department. During the past 6 years, they have provided the necessary support, guidance, and assistance to help me pursue my degrees here at RIT. Additionally, I would like to thank my committee members, Dr. Jeff Mills and Dr. David Ross, for providing support and expertise in my experimentation and analysis. Their input saved countless headaches and hours of unproductive work.

In addition to my committee, I would like to extend a big thank you to my research team: Julia Faraone, Olivia Fraser, and Symeon Bushunow. Their assistance, cooperation, and reliability were crucial to the success of this project.

Having two committee advisors certainly provided many valuable aspects to my project. Firstly, I would like to thank Dr. George Thurston for everything he has done for me in the past 4 years. Always quick with a joke, and ready to answer any questions I had, Dr. Thurston's assistance certainly made my time spent in his lab enjoyable and unforgettable.

My final thank you is to my other research advisor, Dr. Lea Michel. Her guidance, support, and constant availability has directly shaped me into the researcher I have become. In addition, her forward looking mentality helped me immensely when it came to making choices about my future career path. Without all of these people, this project would have never come to fruition.

III) ABBREVIATIONS

APS	Ammonium Persulfate
AIC	Akaike Information Criterion
$CaCl_2$	Calcium Chloride
CH_3COOH	Glacial Acetic Acid
D_2O	Deuterium Oxide
DOSY	Diffusion Ordered Spectroscopy
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FF	Fast Flow
FID	Free Induction Decay
FPLC	Fast Protein Liquid Chromatography
HCl	Hydrochloric Acid
HR	High Resolution
HSQC	Heteronuclear single-quantum coherence
IOL	Intraocular lens
IEC	Ion Exchange Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
KH_2PO_4	Monopotassium Phosphate
LB	Lysogeny Broth
$MgSO_4$	Magnesium Sulfate
Na_2HPO_4	Disodium Phosphate
NaCl	Sodium Chloride
NaH_2PO_4	Monosodium Phosphate
NaOH	Sodium Hydroxide
NH_4Cl	Ammonium Chloride
NMR	Nuclear Magnetic Resonance
OD	Optical Density
PAGE	Polyacrylamide Gel Electrophoresis
PCS	Posterior subcapsular cataracts
PDB	Protein Data Bank
SDS	Sodium Dodecyl Sulfate
SEC	Size Exclusion Chromatography

TEMED	Tetramethylethylenediamine
WHO	World Health Organization
WT	Wild Type

IV) Table of Contents

I) ABSTRACT: ***III***

II) ACKNOWLEDGEMENTS ***IV***

III) ABBREVIATIONS ***V***

IV) Table of Contents ***VII***

1 INTRODUCTION ***1***

1.1 Human Eye **1**

1.1.1 Cataracts 1

1.1.2 Statistics 3

1.1.3 Causes 4

1.1.4 Treatment 5

1.2 Crystallin Proteins **7**

1.2.1 α-Crystallin 8

1.2.2 β/γ-Crystallins 9

1.3 Using NMR Spectroscopy to Probe Intermolecular Interactions Between Crystallins **11**

1.3.1 What is NMR? 12

1.4 Brownian Dynamics **16**

1.4.1 Rotational Diffusion 16

1.4.2 T_1/T_2 Experiments 20

1.4.3 Translational Diffusion 24

1.4.4 DOSY Experiment 24

1.4.5 DOSY Gradient Calibration 26

1.4.6 Durbin-Watson Statistical Analysis 26

1.4.7 Akaike Information Criterion (AIC) 27

2 MATERIALS AND METHODS 28

2.1 Preparation of Buffers: 28

2.1.1 Phosphate Buffer: 28

2.1.2 Sodium Hydroxide Wash Solution: 28

2.1.3 Sodium Acetate buffer: 28

2.1.4 NMR Buffer 29

2.2 Protein Expression 29

2.3 Sonication 30

2.4 Concentrating 30

2.5 Protein Purification 30

2.5.1 Size exclusion Chromatography using Pharmacia Biotech P-50 Protein Pump 31

2.5.2 Size exclusion Chromatography using BIO-RAD NGC Chromatography Pump 31

2.5.3 Cation Exchange Chromatography using Pharmacia Biotech P-50 Protein Pump 32

2.5.4 Cation Exchange Chromatography using BIO-RAD NGC Chromatography Pump 33

2.6 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) 34

2.7 Buffer Exchange for Ion Exchange Chromatography 35

2.8 Concentration via Centrifugation 35

2.9 UV Spectroscopy 35

2.10 NMR Spectroscopy 36

2.10.1 T_1/T_2 HSQC 36

2.10.2 NMR Sample Preparations and Dilutions 37

2.11 Calibration Standard 38

2.12 Calculations 38

3 Results 39

3.1 SDS PAGE Gel 39

3.2 UV Spectroscopy 41

3.3 T_1/T_2 NMR Results 42

3.3.1 τ_c Times, Single Exponential Analysis 42

3.3.2 (T_1) Double Exponential Analysis 44

3.3.3 τ_c Times Double Exponential Analysis 46

3.3.4 Comparison of Varian Approximation with Full τ_c Formula 47

3.3.5 (T_1) Time Constants with a Single Exponential Decay 49

3.3.6 (T_1) Time Constant 1 from Double Exponential Decay Analysis 50

3.3.7 (T_1) Time Constant 2 with a Double Exponential Decay 51

3.3.8 T_2 Time Constant 52

3.3.9 Hydrodynamic Radii 53

3.3.10 Durbin Watson Single Vs. Double Exponential Analysis 33 mg/ml 54

3.4 Akaike Information Criterion Results 56

3.5 Translational Diffusion Coefficient Measurements 57

4 Discussion 62

4.1 Purification Process 62

4.2 Rotational Correlation (τ_c) 63

4.3 Single Vs. Double Exponential Analysis 64

4.4 Gradient Calibration 68

4.5 Translational Diffusion 69

5 Conclusion 72

6 References 75

7 Supplemental 81

7.1 Other τ_c Values recorded over the past 4 years 81

7.2 Time Constants 82

7.2.1 T_1 Single Exponential Time Constants 82

7.2.2 T_1 Double Exponential Time Constants 82

7.3 Durbin Watson Analysis 83

7.3.1 Single Exponential Decay 83

7.3.2 Double Exponential Decay 83

7.4 Equations 84

7.4.1 Hydrodynamic Radius Approximation: 84

7.4.2 τ_c Approximation used on Varian Software 84

7.4.3 Stejskal-Tanner Equation 84

7.5 Spectra 86

7.5.1 DOSY Spectrum of Sucrose at 2% Gradient Strength 86

7.5.2 DOSY Spectrum of 33 mg/ml sample of γB-crystallin at 2% Gradient Strength 87

1 INTRODUCTION

1.1 Human Eye

Transparency in the eye is due to the short-range interactions of proteins in the Crystallin family. Under healthy conditions, light travels through the iris before being focused by the lens. The focused light lands on the optical nerve, which then transfers the signal to the brain. The human eye lens grows throughout the entirety of a lifespan, although it slows in growth rate with age. In mammals, lens epithelial cells are the precursors to lens fiber cells.[1] The lens fiber cells are filled with the crystallin proteins that are responsible for transparency of the lens.

The lens must maintain a certain concentration of crystallin proteins to facilitate healthy short-range interactions. Light-scattering, an analytical method used to measure how light particles deviate from a straight trajectory, has been performed on a mixture of the crystallin proteins. The results of these experiments suggest an interesting relationship between light scattering and protein concentration. Initially, as concentration is increased, an increase of light scattering is observed, as expected. But, once the protein concentration reaches 0.2 g/ml, a decrease in scattered light is observed, which accounts for the observed transparency. This is due to the short range, liquid-like spatial ordering of the proteins, similar to that of glass.[2]

1.1.1 Cataracts

Cataracts are the leading cause of blindness in the world.[3] In the United States, cataracts are the leading cause of vision loss.[3] Cataracts are defined as “any opacity of the crystalline lens.”[4] Cataracts in the eye occur when the proteins in the cytoplasm aggregate or phase separate.

When the lens becomes clouded, the light which travels through the eye lens is scattered. The result of this is reduced vision, and ultimately blindness if left untreated.

There are three main types of age-related cataracts: Nuclear Sclerotic, Cortical, and Posterior Subcapsular. As mammals age, any one type, or combination of the three, may develop.

A *nuclear sclerotic cataract* generally forms slowly over many years and is one of the leading causes of vision loss in the elderly. As the lens ages, the lens nucleus becomes compressed and hardened as a result of a buildup of new layers of lens fibers. This results in the sclerotic lens nuclei decreasing in transparency and can result in slight visual aberrations, as well as a nighttime glare. Generally, these cataracts do not have an overly detrimental effect on vision, but as they become more severe, sufferers may note a loss of color discrimination, as well as a decrease in far-sighted vision. As the cataract matures, the material within the cortex begins to liquify, resulting in opacity within the nucleus. If left untreated, cortical material may leak across the capsular lining, resulting in an inflammatory response.

The cortex of the lens is comprised of the newest lens fibers, and lens fibers are not lost with age. The new lens fibers build up on the outside of the lens, just below the capsule of the lens. As a mammal ages, cortical spokes, a uniquely star-shaped type of opacity, may develop within the cortex of the lens. Generally, this does not cause any vision loss, unless the spokes involve the visual axis or cover the entire cortex. When this happens, the lens becomes visibly white and cloudy, resulting in a *cortical cataract*.

Posterior subcapsular cataracts (PCS) are the result of posterior migration of lens epithelial cells responding to an external stimulus. Usually this is a spontaneous reaction, but PCS may also be brought on as a result of certain metabolic causes including diabetes,

inflammation, uveitis, and long-term use of topical corticosteroid use. This type of cataract is characterized by granular opacities located in the central posterior cortex, is most commonly associated with younger patients, and usually results in near-sightedness.

1.1.2 Statistics

As of October 2018, the World Health Organization estimated that over 1.3 billion individuals were suffering visual impairment due to cataracts.[5] Numerous large scale population based studies have been conducted, the results of which strongly suggest that the prevalence of cataracts increases with age. The likelihood of a cataract increases from 3.9% at ages 55-64 to 92.6% by age 80.[6] By age 75, 50% of white Americans will have a cataract, and by age 80, this number jumps to 70%. Cataracts can affect anyone regardless of age, ethnicity, or gender, although recent studies suggest Caucasian individuals have a higher prevalence of and earlier onset age for cataracts compared to people of other races.[7]

Cataracts can affect both men and women, although studies suggest women are more susceptible to cataracts then men.[7] Additionally, it has been suggested that the presence of cataracts may be associated with an increased mortality rate. This association may be caused by the not yet understood link between cataracts and conditions such as type-2 diabetes and smoking.[8]

As the U.S. population increases, the median age of a U.S. citizen is expected to increase from 38 in 2019 to 43 by 2060.[9] Because of the correlation between cataracts and age, the number of individuals affected by cataracts is expected to grow as the elderly population increases in size. The National Eye Institute estimates that the number of individuals suffering from with cataracts will grow from ~20 million in 2010 to 50 million by 2050. Between 2000

and 2010, the number of cataract cases in the U.S. rose by 20%, and this number in expected to rise more quickly as the population ages.[7]

1.1.3 Causes

There are numerous causes of mammalian cataracts, including injury, aging, genetic disposition, and ultraviolet light exposure. Additionally, certain health problems such as diabetes, obesity, hypertension, and smoking may increase the chances of developing cataracts, in ways that are not yet fully understood.

Ocular trauma is one of the leading causes of cataracts in mammalian lenses. Approximately one in five individuals will experience some degree of ocular trauma in their lifetime, although just 2% of population will experience a trauma severe enough to warrant hospitalization.[10] Portions of the population who experience repeated ocular trauma, such as boxers, are much more likely to develop traumatic cataracts than the general population.[11] When a blunt trauma occurs, generally a stellate or rosette shaped posterior axial opacity will form. The timeframe for the formation of a traumatic cataract is very acute, and once formed, the cataract usually remains stable over time. When a penetrating trauma occurs, it is generally accompanied by a disruption of the lens capsule. This leads to cortical changes that may remain localized or may progress rapidly to total cortical opacification. In the case of minor ocular trauma, when the lens capsule is not damaged, cataracts often form long after the injury as a result of a disturbance to the growth of new lens fibers from epithelial damage.[12] Proper care and medical attention, along with early detection, are essential for diagnosis and prevention of ocular trauma cataracts. [13]

Some individuals are genetically disposed to cataract formation. Often, cataracts are caused by a point mutation in one of the proteins within the eye. A point mutation occurs when a single amino acid is replaced with a different amino acid.

Numerous individual mutations in the human γD crystallin protein have been linked to the development of early onset cataract disease, including R14C, R58H, and R36S. Many of these point mutations lead to the formation of either covalently linked aggregates (R14C) or crystal formation (R58H and R36S).[14] Another interesting mutation, P23T or P23V, can lead to cataract formation by drastically reducing the solubility of the crystallin protein, while simultaneously allowing the protein to maintain its native fold. [14,15] Understanding how a single point mutation can lead to aggregation or phase separation is an area of great interest to researchers, since it could lead to the development of new cataract treatment or prevention methods.

Numerous studies have identified ultraviolet (UV) light exposure as a risk factor for cataract formation.[16] These studies suggest that individuals with greater exposure to UV light correlates with a slightly increased risk of cataract formation. Additionally, tests on laboratory animals suggest that exposure to artificial UV light sources leads to opacities forming within the eye.[16] The mechanisms in which UV light induces cataract formation are not fully understood, but studies suggest that they may be caused by either tryptophan degradation, which leads to decreased levels of ATPase and increased levels of free radicals, or UVB damage to the lens epithelium which leads to an imbalance of calcium within the lens membrane.[17]

1.1.4 Treatment

Presently, surgical removal of a cataract is the standard method of treatment. The first type of surgical treatment of cataracts is called intracapsular cataract extraction (ICCE). This method involves the complete removal of the lens, leaving the patient in an aphakic state. With the lens completely removed, the patient is forced to wear thick lensed glasses, which often result in permanently distorted peripheral vision. ICCE was the prominent treatment option until the mid 1970's, at which point ophthalmologists began using an improved method involving the

use of intraocular lenses (IOL's). Intraocular lenses are "polymeric devices implanted in the globe of the eye and intended to replace the cloudy, cataractous natural lens."[18] Essentially, this method involves the removal of the natural lens, and replacement with the IOL.

Since the 1970's, success rates for cataract treatments have steadily increased due to technological advancements. Cataract surgery first requires the removal of the natural lens. During cataract surgery, minimizing damage to other the parts of the eye is crucial. After local anesthesia is administered, a small incision must be made in which the clouded lens is removed, and the new lens inserted. It is essential to remove as much of the old lens as possible for optimal post-surgery results, as leaving small amounts of the clouded lens behind may lead to regrowth of the cataract. One technological advancement that has greatly improved cataract surgery success rates is the utilization of phacoemulsification, a method that employs ultrasonic waves to emulsify the old lens in the eye.[19] This method has allowed surgeons to make much smaller incisions in the eye, thereby reducing the potential of post-surgical complications and reducing recovery time.

In developed countries, the success rate for cataract corrective surgery is as high as 95%, with fewer then 5% of surgeries involving complications like inflammation or infection. In 2015, 3.4 million corrective surgeries were performed in the United States alone, while an additional 16.6 million were performed across the globe.[20] Although cataract treatments are generally considered successful and routine, the cost of treatment is of great concern. In 2017, the US spent $10.7 billion treating cataracts.[21] While cataract corrective surgery is generally covered under health insurance, the cost of surgery often forces people who lack insurance to live with the symptoms associated with cataracts rather than seek treatment. Additionally, in developing countries, cataract treatment encounters issues that are not seen in developed countries, including

a shortage of trained medical personal, lack of hygienic conditions, and most notably, cost. A crude treatment known as “couching”, which involves using a needle to push the clouded lens to the back of the eye, is still commonly used in countries with less access to medical treatment.[22,23] Couching was the first documented treatment for cataracts and was utilized by the Egyptians back in 2600 BC.[24] An updated method described as “manual sutureless small incision extracapsular cataract surgery” (MSSIECS) has been proposed as a promising treatment for patients in developing countries.[25]

Although modern cataract treatment methods show high success rates, there is still the risk of complications. One such complication, known as a “secondary cataract,” occurs when the surgeon fails to fully remove all of the lens cells from the eye. Over time, the eye begins to develop a secondary cataract in response to the wound created from the surgery and the remaining epithelial cells attempt to heal the eye by producing a mixture of scar tissue and new epithelial cells.

Future treatment goals of cataracts will focus on the protein misfolding aspect of cataracts. Since cataracts are often caused by misfolded proteins, understanding what causes the misfolding may lead to development of better treatment and preventative measures. However, understanding the source/cause of crystallin protein misfolding and aggregation is an extremely complex problem. A better understanding of how the crystallins behave in solution may begin to provide the necessary framework for the development of models that can be used to better understand the protein-protein interactions that lead to cataracts.

1.2 Crystallin Proteins

The Crystallin family of proteins was first described by C.T. Mörner in 1893.[26] Crystallin proteins account for approximately 90% of the water-soluble protein within the lens and ~35% of

the mass of the lens. The Crystallin family of proteins is split into two distinct gene families: the α-crystallins and the βγ-crystallins.[1] Familial distinctions are characterized by their genetic organization, the regulation of their expression pattern, and their role in numerous diseases. Although all of the crystallin proteins are found in the mammalian lens, experimentation has shown that relative proportions and concentrations of α, β, and γ vary throughout the eye.[28] When originally discovered, crystallins were thought to fill the large elongated, terminally differentiated fiber cells of the lens. This role ensured adequate transparency and molecular organization required to maintain the necessary refractive properties of the lens. But recent studies have shown that crystallin proteins can be found throughout the body, performing a variety of important biological functions.

1.2.1 α-Crystallin

The α-crystallin complexes comprise 40% of the total proteins in the lens and are the largest of the Crystallin family, with a mass on the order of 600-900kDa in mammals.[29] The complexes are composed of two subunits, αA- and αB-crystallins, that share 57% sequence identity and exist within the lens at a molar ratio of roughly 3:1.[29] The main roles of the α-crystallins are to act as molecular chaperones and to prevent aberrant protein interactions.[30] Part of the α-crystallins' chaperone-like properties include the ability to prevent the precipitation of denatured proteins and to increase cellular tolerance of stress. Due to their role in preventing protein misfolding, it is believed that the chaperone tendencies of the α-crystallins are responsible for the maintenance of long-term lens transparency by preventing non-specific aggregation of other crystallins.[29] Additionally, in 1989, it was first discovered that α-crystallin could be found also in the heart, skeletal muscle, skin, brain, spinal cord, and lungs.[31] The α-crystallins belong to the small heat-shock protein family.

1.2.2 β/γ-Crystallins

Originally, the β/γ family of crystallins was split into two separate families based on varying isoelectric points and aggregation properties. However, now, they are usually grouped together into a single family due to their shared motif of 4 antiparallel β sheets and homologous amino acid sequences.[32] Generally, β-crystallins exist as oligomers, while the γ-crystallins are neatly folded, globular monomers.

The γ-crystallins are the earliest of all crystallins to be expressed and are the most concentrated within the nucleus; the β-crystallins are the second most concentrated protein component. The γ-crystallins occur in the mammalian lens at concentrations of > 400 mg/ml, and are the smallest of the crystallin family, comprised of approximately 175 amino acids and a mass of 21kDa. Mammalian genomes contain 7 different γ-crystallin genes. Of these, 6 of them (γA-γF) are closely related, linked by a tandemly repeated gene cluster and highly similar amino acid sequence. The 7th γ-crystallin, γS, is located on a separate chromosome and has a more divergent sequence than the other 6. Although 7 different γ-crystallins have been discovered, only γC and γD are found in the lens (γE and γD are pseudogenes).[33] Because of the high concentration of γ-crystallins, they are largely responsible for the maintenance of the transparency of the mammalian lens. Because of this, γ-crystallins are of particular interest with regard to understanding cataracts, especially since there are over 30 known mutations that have been proven to lead to congenital cataracts.[33]

Taking deeper look at γ-crystallin structure is necessary to understand how they dictate lens transparency. The Greek key motif is a secondary structural element that is comprised of 4 antiparallel β sheets and is common to all γ-crystallins. The γβ-crystallin proteins have two domains, with each domain containing 2 of the Greek key motif secondary structures. This motif

has numerous favorable aspects and properties suited to their role in maintaining transparency within the lens. Firstly, the motif allows for tight intramolecular packing, which is required to maintain a concentration of >400 mg/ml and creates a neat monomeric globule of approximately 5nm in size, with a Stokes radius of 2.13 nm. Tight packing gives γ-crystallins a very low frictional ratio of 1.21, which is only slightly above that of a perfect sphere (1.12).[34] This structural characteristic is thought to contribute to its low propensity for interaction with both solvents and other proteins.

There are some important key characteristics common to all of the crystallin proteins, including their ability to form stable and durable structures and their enhanced solubility. The γ-crystallins are known to be extremely stable proteins, with melting points ~80°C and a robust resistance against denaturation via urea and guanidinium chloride.[35] The origin of this stability is the subject of intensive research, although not yet fully understood. It has been proposed that the tightly packed nature of the double Greek key plays a large role in its stability. This argument only partially explains the stability of the γ-crystallins, because the β-crystallins share this same motif, yet display significantly lower stability. The overall stability of the γ-crystallins is likely enhanced by a mixture of factors including H-bonds, van der waals packing, the hydrophobic effect, aromatic stacking, ion pairs, and salt bridges. The stability of the γ-crystallin is an important component to its ability to control the transmittance of light in the lens. As the proteins lose their thermodynamic stability, large aggregates can form, increasing the amount of light scattering and decreasing vision.

The bovine γB protein is an excellent candidate for laboratory studies, as it is a homolog to the human γD protein. The bovine γB protein (Protein Data Bank ID 1AMM) is 174-residues with a molecular mass of 20,992.56 Da, while the human γD protein (PDB 1HK0) is 173-

residues with a mass of 20,634.97 Da. Using UCSF Chimera, a sequence and structural alignment was performed.[36] The two proteins share a 75.29% sequence identity, composed of 133 residues, with a root mean square deviation (RMSD) value of 0.904Å between the 173 aligned backbone residues, suggesting that the two homologs are highly similar in structure. Figure (1) shows an overlay of the two homologs for structural comparison. Bovine γB-crystallins are often studied in experiments, because of the vast amount of previous experimental data already collected on them (NMR assignments, phase diagrams, and purification methods) and their similarity to the human γD protein.

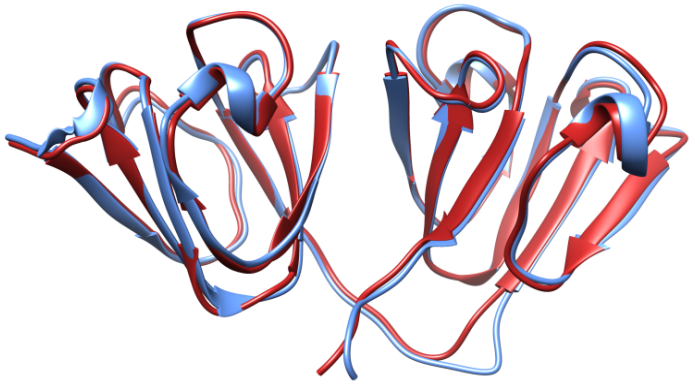


Figure 1 Structural alignment of human-γD and bovine γB-crystallins

An overlay of human γD-crystallin (PDB 1HK0) in red and the bovine γB-crystallin (PDB 1AMM) in blue. Structural comparison performed in UCSF Chimera gave a RMSD of 0.904Å between 173 aligned backbone residues.

1.3 Using NMR Spectroscopy to Probe Intermolecular Interactions Between Crystallins

Nuclear Magnetic Resonance (NMR) is an analytical technique that utilizes the magnetic properties of atomic nuclei to provide chemical and structural information about molecules.

NMR is both a qualitative and quantitative technique and can be used to study a variety of substances ranging from simple organic molecules to complex proteins. This powerful analytical method can even be used for the analysis of counterfeit food and drug imports.

1.3.1 What is NMR?

To understand how NMR works, a basic understanding of the relationship between atoms, nuclei, and molecules is necessary. Every atom is constructed from a positively charged nucleus, consisting of positively charged protons and uncharged neutrons. The nucleus is surrounded by a negatively charged field of electrons. A molecule is a collection of two or more atoms bound together in a particular molecular structure. NMR is an extremely powerful tool that can be used to determine those molecular structures and dynamics between molecules

When a molecule is exposed to a magnetic field, each atom will feel a marginally modified field as a result of the magnetic shielding effects caused by neighboring electric charges and the nuclei and electrons of neighboring atoms. Therefore, the magnetic field each atom experiences depends on its surrounding environment. NMR takes advantage of these small fluctuations in response to an applied magnetic field with the use of an extremely sensitive detector.

NMR requires that the molecule of interest contains atoms with an odd number of protons. All elements with an odd number of protons exhibit a property called spin. Only isotopes that have non-zero spin are detectable by NMR- these elements are defined as “NMR-Active.” Since all nuclei are electrically charged, as they rotate, they create their own individual magnetic field, known as the magnetic moment (μ). All of the magnetic moments in a molecule have a magnitude and direction. In general, in the absence of any external magnetic fields, these

magnetic moments will all point in "random" directions and therefore add to zero. This is known as degeneracy.

In NMR, a strong external magnetic field, usually denoted as B_0, will influence the individual spin states of a molecule. Nuclei such as ^{1}H, ^{13}C, ^{15}N, and ^{31}P have ½ spins. In a large applied magnetic field, approximately half of the spins will point in the same direction as the applied field and approximately half of the spins will point in the opposite direction as the applied field. The Boltzmann's distribution dictates that there will always be more nuclei with low energy spin states (that align with the field) than high energy spin states (not aligned with the field). This phenomenon leads to a population excess known as bulk magnetization (M). The difference between the two spin energy levels is ΔE, which increases as the applied field strength increases. ΔE is represented by (Eq. 1), where h represents Planck's constant, γ is equal to the gyromagnetic ratio, and B_0 is the strength of the applied magnetic field.

$$\Delta E = \frac{h\gamma B_0}{2\pi} \quad (1)$$

An NMR spectrum is the result of pulsing varying frequencies of RF radiation into the sample. When the energy of the applied RF radiation matches ΔE, the nuclei will absorb the energy. Since energy is directly proportional to frequency based on the Planck-Einstein relation, E=hν, this means only a specific frequency of energy can be absorbed by the nucleus. The frequency of the RF radiation that is absorbed by the nucleus induces resonance. The specific frequency of RF radiation absorbed by a nucleus is called the Larmor frequency, (ν), and is

calculated by (Eq. 2), where B_0 is equal to the applied magnetic field, and γ is the gyromagnetic ratio. We can determine the Larmor frequencies of NMR active atoms by observing where "peaks" appear in the spectrum. Using (Eq. 1), we can calculate the amount of RF radiation required to match ΔE.

$$\nu = \frac{\gamma B_0}{2\pi} \tag{2}$$

When the correct energy of RF radiation is pulsed onto a set of nuclei, the energy can be absorbed, resulting in excitation of one of the nuclei from the low to the high energy state. This can be done repeatedly, until an even distribution between spin states occurs, a condition known as saturation. At this point, a spin has an equal chance of dropping to a lower energy level as it does of being promoted to a higher energy level. Because at equilibrium there will always be more nuclei in the lower energy state, once the RF radiation pulses stop, the system will relax back to the Boltzmann equilibrium through a phenomenon known as spin-lattice relaxation, also known as T_1 relaxation.

The more spin flips that occur, the stronger the NMR signal, but after a certain point, saturation becomes an issue. To circumvent this, delays are placed between repeated RF pulses, which work together to yield an average signal. A series of pulses and relaxations coupled with acquisition times is known a pulse sequence. The more acquisitions performed, the better the signal to noise ratio, allowing smaller peaks to be discerned from the noise.

Precession is another important aspect of NMR theory. Precession is used to describe how the magnetic moment of a nucleus "wobbles" in a circle around the applied magnetic field,

B_0, at its resonance frequency. Each of the individual moments precess at their resonance frequency, but in a disordered way. Averaging all of these signals provides us with Bulk Magnetization (M), which aligns with the applied magnetic field. In NMR, magnetization is measured in the xy plane. If the bulk magnetization is in the z plane, no signal will be detected. To detect a signal, radio frequency pulses are used to knock the bulk magnetization out of the z axis and into the xy plane. This induces an ordered precession of the bulk magnetization, which then dephases over time through relaxation known as T_2 relaxation. While precessing in the xy plane, the resulting oscillating magnetic moment can be measured in the y axis and will alternate between a highly positive and highly negative value, decaying over time. This is the NMR signal measured by the spectrometer and is a measurement of amplitude and frequency over time, also known as the free induction decay (FID). Once the FID is collected, a Fourier transformation is applied to the FID, which converts the data from the time to the frequency domain. The Fourier transformation is able to distinguish and separate individual signals within the FID, and results in a typical NMR spectrum with signals or "peaks" at corresponding frequencies.

NMR spectroscopy is an extremely sensitive technique. As stated above, nuclei of the same isotope experiencing the same applied magnetic field B_0 will share identical resonance frequencies. However, nuclei are surrounded by electrons, which produce a small but measurable magnetic field that opposes the applied magnetic field B_0. This small amount of opposing magnetic field has a "shielding" effect on the nucleus, meaning the magnetic field the nucleus experiences is actually slightly weaker than B_0. Varying electron densities will produce small variations in the magnetic field, and therefore cause slight variations in the resonance frequencies of specific nuclei.

1.4 Brownian Dynamics

Brownian motion is an umbrella-term used to describe the random erratic motion of microscopic particles within a liquid solution. These motions are governed by the kinetic molecular theory, in which the directions of the molecules' motions are random, and the speed at which they move is dictated by the temperature of the system.

Molecular diffusion is a more specific theory that uses Fick's law to describe how molecules will move from higher to lower areas of concentration. Molecular diffusion is actually a consequential result of Brownian motion. Molecular diffusion is described in the simplest terms as the "thermal motion of all particles at temperatures above absolute zero." The rate of diffusion can be affected by a multitude of factors including temperature, concentration, viscosity, and particle size. Molecular diffusion is an important factor of cataract formation, as diffusion measurements experimentally determined for the crystallin proteins may provide insight into aggregation and phase separation rates under varying conditions.

1.4.1 Rotational Diffusion

Rotational diffusion is "the process by which the equilibrium statistical distribution of the overall orientation of molecules or particles is maintained or restored."[37] A point of interest regarding the rotational diffusion of a molecule is a characteristic time constant called the rotational correlation time (τ_c). τ_c is the average time that it takes for a particle to rotate one full radian. This value is important, because it can provide insight into the average size of the rotating particles in solution.

To understand the process by which we measure rotational diffusion, one must first understand the mechanisms in which relaxation occurs for nuclear spins. For spin ½ nuclei,

relaxation of the spins is the direct result of fluctuating magnetic fields at the particular site of the spin. These fluctuating magnetic fields are caused by the thermal motion of the molecules. For example, as a molecule tumbles through solution, both the magnitude and the direction of the magnetic field created by a particular spin felt by its neighbor will change as a result of the motion of the two spins. If one monitors individual signals in the transverse field at a specific frequency, this will produce fluctuations in the signal specific to each spin. Because the fluctuations for individual spins are dependent on each other, they will generally share the same timescale and amplitude. Fields of neighboring magnetic dipoles in an isotropic liquid are such that the average amplitude of the relevant fluctuating fields is zero. To understand the impact of the fluctuations, we can employ an *autocorrelation function*. Fast fluctuations will cause the autocorrelation to diminish quickly, while slower fluctuations result in a slower decay of the *autocorrelation function*. The general theory behind calculating the *autocorrelation function* involves the comparison of the field at a given time, t, with a later point, t+τ. By comparing the interval, τ, to the time scale of the fluctuations within the field, the *autocorrelation function* for a sphere can be calculated:

$$G(\tau) = (B_x^2)e^{-|\tau|/\tau_c} \quad (3)$$

Where $B^2{}_\chi$ represents the mean square fluctuating field, τ equals the time interval, and τ_c is the *autocorrelation time* of the fluctuations. This value represents the isotropic diffusion of a rigid rotor. The *autocorrelation function* is a monotonically decreasing function of τ_c. Rapid fluctuations produce smaller τ_c values, while slower fluctuations result in larger τ_c values.

To further understand the fully derived equation to solve for τ_c, one must examine the spectral density J(ω). This value is the probability function used to identify motions at a given angular frequency, ω and is defined as twice the one-sided Fourier transform of the autocorrelation function. The relevant spectral density for our NMR analysis corresponding to the isotropic diffusion of a sphere, J(ω) is defined as:

$$J(\omega) = \frac{2}{5}\frac{\tau_C}{(1+\omega^2\tau_c^2)} \tag{4}$$

This function represents the spectral density formula for spheres, taking into consideration τ_c, ω, and correlation time. The equation summarizes the noise power as a function of frequency, which results from rotational diffusion.

For spin ½ nuclei, there are multiple relaxation mechanisms, including dipole-dipole, J-coupling, quadrupolar, and chemical exchange. However, quantitatively, dipole-dipole relaxation has the greatest influence on overall relaxation. The dipole-dipole relaxation rate is proportional to the r^{-6} of the separation between the two nuclear dipoles, and thus is highly sensitive to distance. To describe these various quantum mechanical transitions, the product operator formalism is used, specifically an IS system, where 'I' corresponds to the proton attached to a nitrogen, with spin, 'S' To calculate the longitudinal relaxation, or T_1 aspect of the spectral density, (Eq. 5) can be used.

$$R_{1s} = \left(\frac{d_{00}}{4}\right)\{J(\omega_I - \omega_s) + 3J(\omega_s) + 6J(\omega_I + \omega_s)\} = \frac{1}{T_1} \tag{5}$$

For spin-spin relaxation, or the T_2 aspect of the spectral density, (Eq. 6) can be used:

$$R_{2S} = \left(\frac{d_{00}}{8}\right)\{4J(0) + J(\omega_I - \omega_s) + 3J(\omega_s) + 6J(\omega_I) + 6J(\omega_I + \omega_s)\} = \frac{1}{T_2} \quad (6)$$

To solve the relevant spectral density, $J_{(\omega)}$, in (Eq. 4), the appropriate gyromagnetic ratios and Larmor frequencies found in table (1) are necessary:

Table 1 Relevant Constants

Proton ^{1}H Gyromagnetic Ratio (γ)	2.67522x10^8 rads / s T
Nitrogen ^{15}N Gyromagnetic Ratio (γ)	-2.7126x10^7 rads / s T
Vacuum Magnetic Permeability (μ_0)	1.25664x10^{-4} Kg M/Ampere^{2}Second^2
RS Separation (R_{IS})	1.023 Å [38]
Planck Constant (h)	6.62607004x10^{-34} m^2 Kg/S

The Larmor frequency is a function of the applied magnetic field strength. In our case, we used a 600MHz magnet, so the values reported are for this field strength and can be adjusted accordingly for a different magnetic field strength.

Both (Eq. 5 & 6) share a common factor, d_{00} which we will call the prefactor. This value is given by:

$$d_{00} = (\frac{\mu_0}{4\pi})^2 h^2 \gamma_I^2 \gamma_S^2 r_{IS}^{-6} \quad (7)$$

Using the values found in Table (1), the prefactor can be calculated for specific field strengths. Once the appropriate spectral densities are determined for I or S, the values can be substituted into (Eq. 5 & 6), along with the experimentally determined T_1 or T_2 values, to solve for τ_c.

Due to the nature of (Eq. 5 & 6), τ_c values can theoretically be calculated using either T_1 or T_2 independently. However, using a quotient of T_1/T_2 allows for an exact cancelation of the prefactor and $R_{IS,}$ which is the distance between the backbone nitrogen and proton in Ångstroms. Additionally, using both values provides a useful check by comparing τ_c calculated from T_1 and T_2 independently.

However, the τ_c value can also be ***approximated*** using a ratio of the T_1 (longitudinal) and T_2 (transverse) relaxation times of the protein in solution (Eq. 8):

$$\tau_c \approx \frac{1}{4\pi\nu_N}\sqrt{6\frac{T_1}{T_2}-7} \tag{8}$$

A comparison of the approximated and fully derived equations for τ_c shows that the approximation is adequate, especially when τ_c is above $1x10^{-9}$ seconds, as seen in (Results 7.3.3).

Once τ_c is determined, that value can be used to calculate the effective hydrodynamic radius of the rotating molecule, as seen in (Eq. 3.3.9). This is a convenient method for better understanding aggregation rates, since the radii of the molecules increase with association.

1.4.2 T_1/T_2 Experiments

T_1 and T_2 are decay constants associated with two different types of relaxation. As described earlier, a magnetic moment (or more commonly referred to as magnetization) knocked

over into the transverse plane will relax back to the z-axis over time. There are two main mechanisms used to describe how the molecules relax: spin-lattice (T_1) relaxation and spin-spin (T_2) relaxation.

Spin-Lattice (T_1) relaxation is often referred to as longitudinal relaxation and refers to the return of magnetization to the equilibrium in the z-axis, or the direction of the applied magnetic field. To measure T_1 relaxation times, an inversion recovery T_1 experiment is utilized. During this experiment 180° pulse is applied to the sample. This causes the z-component of the net magnetization vectors to rotate into the negative (-) z-axis. After the 180° pulse, the first delay (τ) is applied, at which point relaxation occurs along the longitudinally plane. During this time, magnetization begins to return to its original equilibrium in the z-axis. Then, another 90° pulse is applied after a certain time period (τ); this second pulse rotates the magnetization that has decayed into the z-axis back into the xy-plane, known as transverse magnetization, where a signal can be measured. The entire pulse sequence repeats with different τ values, resulting in different amounts of recovered signal, depending on the molecule's T_1 relaxation time.

As the net spin relaxes back to the Z-axis, the overall energy of the spin system decreases due to the statistical favoring of the lower energy, spin up, and parallel orientation. This means that energy is lost from the system during T_1 relaxation, in the form of heat transfer called *thermal relaxation.* This energy is transferred to surrounding nuclei, through collisions, rotations, and other various electrostatic interactions. The relaxation time, T_1, is used to quantify the rate of transfer of energy from a particular spin system into its neighboring molecules, hence the term *spin lattice.* The actual time constant value for T_1 relaxation can be calculated using the following (Eq. 9):

$$M_z = (M_0 - 1)e^{-\frac{t}{T_1}} + 1 \quad (9)$$

where M_z is the net recovered magnetization after a certain relaxation time, t, and M_0 represents the initial maximum value for the magnetic moment, M. T_1 relaxation is generally explained as a single exponential decay, although our results indicate that a double exponential decay can occur. For this to occur, (Eq. 9) would require that a second population of decay be present, which we will denote as M_{z2}. The value for T_1 sets the theoretical upper limit for possible values of T_2, since $T_2 \leq T_1$. This is because fluctuating molecular fields cannot maintain coherence of the transverse spin polarizations while rotating the individual spin polarizations towards the z-axis. For biological materials, T_1 times are generally on the time scale of 0.05 to 5 seconds.

Spin-Spin (T_2 Relaxation), often referred to as transverse relaxation, is a measure of the decay of the excited net magnetization perpendicular to the applied magnetic field.[39] To measure T_2 relaxation times, 90° pulse is applied to the sample, causing the net magnetization of the sample to fall into the xy-plane. Initially, in the xy-plane, the spins are oscillating in coherence. Over time, the individual spins begin loose coherence, or de-phase with each other, and eventually the signal decays to zero. This is due to local fluctuations in the magnetic field caused by surrounding nuclei. The T_2 time constant is calculated by the function:

$$M_{xy} = M_0 e^{-t/T_2} \quad (10)$$

where M_{xy} is the signal in the xy plane after a certain relaxation time, t, and M_0 is the initial signal in the xy plane. As the magnetic moment begins to de-phase, the individual nuclei begin to cancel each other's signal out, thereby resulting in a reduction of the net magnetization signal in

the xy-plane. T_2 relaxation may occur with or without a contribution from T_1 relaxation and is always faster than T_1 relaxation.

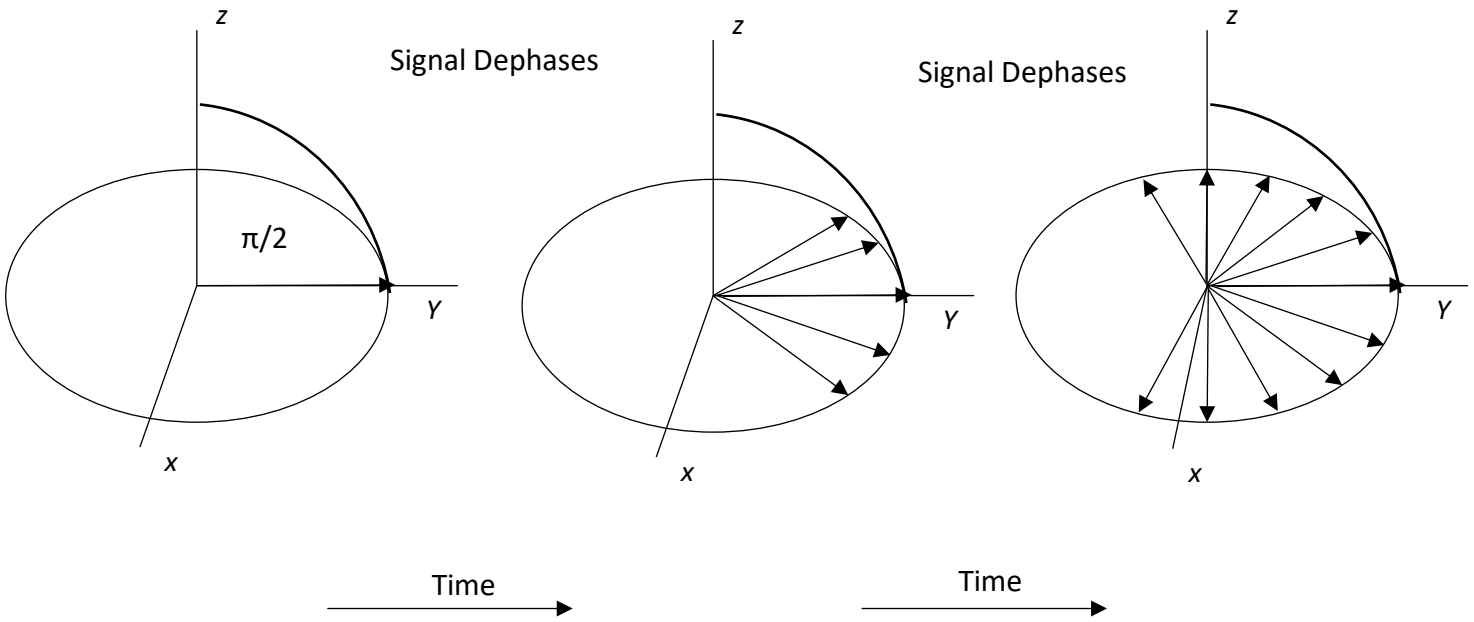


Figure 2 Example of T_2 Relaxation

T_2 Dephasing occurs over time in the XY plane. Signal decreases as net magnetization undergoes dephasing at a rate of 1/e.

T_1/T_2 NMR experiments are convenient techniques to study protein dynamics for variety of reasons. Firstly, they only require ^{15}N-labeled protein, which is less expensive to acquire than ^{13}C-labeled protein. Secondly, both T_1 and T_2 experiments are typically only 1D experiments, so they can be collected quickly and at different temperatures and concentrations.

1.4.3 Translational Diffusion

Translational diffusion is the counterpart to rotational diffusion and describes "the maintenance or restoration of the equilibrium statistical distribution of particles' position in space." The diffusion coefficient is a mathematical value used to describe the ratio between the molar flux due to molecular diffusion and the concentration gradient of a particular species. It is an important value found in numerous physical chemistry laws, including Fick's law, which is used to describe the rate at which molecules diffuse based on size and concentration. The translational diffusion coefficient is affected by numerous factors including, temperature, concentration, and most importantly, viscosity. By experimentally determining the translational diffusion coefficients for the γB-crystallin under different conditions, we can glean important information about the behavior of the protein molecules, and how the kinetics of the molecules change as the proteins associate. Diffusion coefficient measurements can be collected using a variety of analytical methods including NMR, light scattering, and scanning fluorescence microscopy.[40] For protein analysis, NMR is a powerful method for determining diffusion coefficients due to the complexity of protein molecules and the sensitivity of the NMR technique.

1.4.4 DOSY Experiment

The most common NMR technique used to measure diffusion coefficients is called Diffusion Ordered Spectroscopy (DOSY), or Self Diffusion (SD)-NMR. DOSY NMR utilizes radio frequency pulses, routinely used in NMR experiments, along with magnetic field gradients that encode the necessary spatial information. In the simplest form of a DOSY experiment, a

pulsed field gradient spin echo (PGSE) is employed. During a PGSE, the net magnetization is excited with a 90° radio frequency pulse before being uniformly dephased using a magnetic field gradient pulse. After a time of Δ/2, a 180° radio frequency pulse is used to invert the dephased dispersed magnetization. Finally, a second gradient pulse is applied, which refocuses the signal. The second gradient pulse is only capable of refocusing the signals from nuclei that have not moved a significant amount longitudinally within the sample tube. Diffusion within the samples will result in some of the nuclei moving longitudinally to a position where the applied gradient will no longer be able to refocus the magnetization, thereby reducing the intensity of the net signal (Fig. 3).

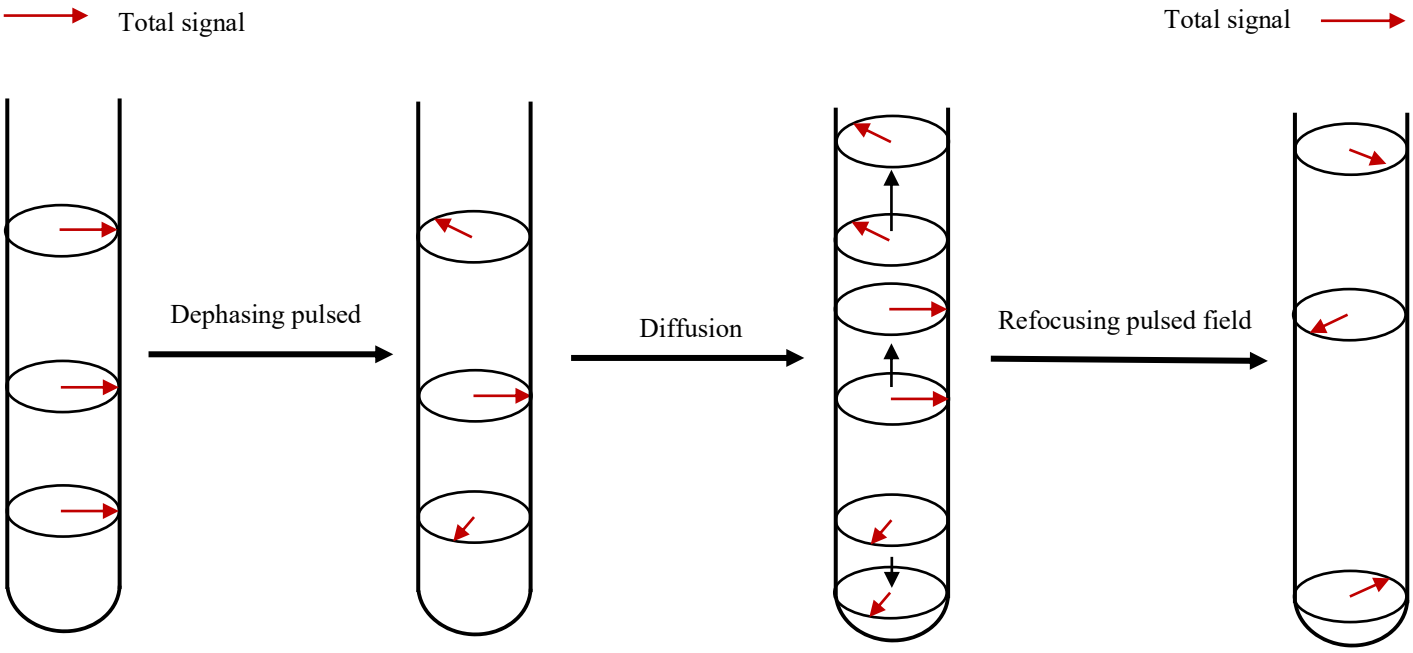


Figure 3 Effect of Diffusion after a Magnetic Field Gradient Pulse

Vertical movement of the molecules, diffusion, results in a reduction of signal after the refocusing pulsed field-gradient.

1.4.5 DOSY Gradient Calibration

Diffusion coefficient values are extremely sensitive to variations in the magnetic field gradient. Due to inherent variations as a result of probe production methods, the strength of the field gradients produced are never exactly determined from the factory.[41] Luckily, using standard solutions, the exact gradient strength can be determined to "calibrate" the probe. To correct, or "calibrate" the gradients, only a single experiment is required. A DOSY measurement is performed on a sample with a known diffusion coefficient. The resulting diffusion data from the calibration standard is fit to a modified version of the Stejskal-Tanner equation, which takes into account the exact gradient strength produced by the probe. Once the calibration has been performed, the correction values are saved into the data file and automatically applied during future DOSY analysis. Common calibration standards include standard doped D_2O samples or sucrose dissolved in 99.99% D_2O.

1.4.6 Durbin-Watson Statistical Analysis

The Durbin-Watson statistical analysis is a test that is used to asses for autocorrelation, or serial correlation in the residuals of a statistical regression analysis.[42] A regression analysis is a statistical method for approximating the correlation between variables, such as time and signal strength. Eq. (11) demonstrates how the Durbin-Watson statistic is calculated:

$$\frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \tag{11}$$

Where e_t represents a given residual, and T is the total number of residuals being tested. A Durbin-Watson analysis will always produce a resulting number between 0 and 4. A Durbin-Watson value of 2 means there is no autocorrelation in the residuals from the statistical analysis. Values between 0 and 2 indicate a positive autocorrelation, while values between 2 and 4 indicates a negative autocorrelation. In this work, the Durbin-Watson test was employed to assess for autocorrelation in the exponential decay functions fit to the T_1 data.

1.4.7 Akaike Information Criterion (AIC)

The Akaike Information Criterion is a mathematical tool that is utilized to test the relative quality of a statistical model applied to a given set of data.[43] It is most often used to compare models for a given set of data, allowing the user to select the best model to use for their analysis. The (AIC) is a general method of evaluating the trade-off between goodness of fit, and the simplicity of a model. Because of this, the AIC can be used as a useful gauge to determine whether a set of data is being overfit. This function works by rewarding for goodness of fit and penalizing by increasing the number of parameters. This is important, because by fitting additional parameters, one can almost always increase the overall fit. Therefore, this penalization is used to reduce overfitting of data models. Once the AIC values are determined for each model, the values of each model's AIC values can be used to determine the relative likelihood of one model fitting a set of data better than another. This makes the AIC function a very useful method in determining how strong a model is, while providing a useful check to combat overfitting. The AIC function can be applied to a model using built-in functions in Mathematica, one of them being corrected for small sample size (AIC); the one used here.

2 MATERIALS AND METHODS

2.1 Preparation of Buffers:

Buffers were prepared using nanopure water (Barnstead NANOpure Diamond), then degassed under vacuum while stirring. All solvents or reagents were either analytical or HPLC grade. When necessary, pH adjustments were made using either concentrated hydrochloric acid (HCl) or concentrated sodium hydroxide (NaOH).

2.1.1 Phosphate Buffer:

Phosphate buffer (50mM) was prepared for use as an elution buffer for size exclusion chromatography at pH 6.8. The following recipe was used to prepare the buffer: 19.3 g of Na_2HPO_4, 8.83 g of NaH_2PO_4, 1.5 g DTT, 1.17 g of EDTA, 46.8 g NaCl, and 0.8 g of Sodium Benzoate in 4 liters of nanopure water.

2.1.2 Sodium Hydroxide Wash Solution:

Wash solution contained 0.5M NaOH and was applied to the size exclusion column in between protein runs. Approximately 40 g of NaOH was added to 2 liters of nanopure water and degassed before use.

2.1.3 Sodium Acetate buffer:

Sodium acetate buffer (0.275 M) was prepared for ion exchange chromatography either with or without NaCl at pH 4.6. Sodium acetate buffer with NaCl (pH 4.6) was prepared with 66.0 g glacial acetic acid (CH_3COOH), 26.8 g NaOH, 0.4 g sodium benzoate, and 38.05 g NaCl in 4 liters of nanopure water. Sodium acetate buffer without NaCl (pH 4.6) was prepared with 66.0 g glacial acetic acid, 0.4 g of sodium benzoate, and 26.8 g of NaOH in 4 liters of nanopure water. Both buffers were made in separate 2-liter containers.

2.1.4 NMR Buffer

NMR buffer was prepared both with and without D_2O. NMR buffer used for sonication was prepared without D_2O. NMR buffer with D_2O was prepared with 0.3168 g NaH_2PO_4, 0.7396 g Na_2HPO_4, 0.2468 g DTT, 72 ml H_2O, 0.016 g sodium benzoate, and 8 ml of D_2O. For NMR buffer without D_2O, the 8 ml of D_2O was replaced with nanopure water.

2.2 Protein Expression

The bovine γB-crystallin protein was expressed in BL-21 Star (DE3) *Escherichia coli* (*E. coli*) cells. The cells contain an IPTG-inducible plasmid with the bovine γB-crystallin gene (no tags). Cells were kept frozen as a glycerol stock solution (8% glycerol) at -80°C. Glycerol stocks and sterile inoculating loops were used to streak cells onto Lysogeny Broth (LB) Agar, which contained kanamycin at a concentration of 100 µg/ml; plates were incubated at 37°C overnight. A single colony was used to inoculate ~25 ml of LB, which contained kanamycin at a concentration of 2.5 µg/ml. Cultures were incubated, shaking at 37°C, 120rpm under aerobic conditions for ~12-18 hours. These small cultures were used to inoculate large growths: 4-liter flasks each containing 200 ml freshly prepared M9 salts (1L nanopure water, 34 g NaH_2PO_4, 15 g KH_2PO_4 and 2.5 g NaCl) with 1.0 g ^{15}N NH_4Cl, and 800 ml of minimal media (2 ml 1M $MgSO_4$, 100 µl 1M $CaCl_2$, 5 g of glucose and 1 ml of 50 mg/ml Kanamycin). These large cultures were incubated, shaking at 37°C, 150rpm until their optical density at 600 nm (OD600) reached ~0.6 (log phase), upon which 0.24g of IPTG was added to each flask. Cultures were induced for an additional 3 hours. The cells were harvested via centrifugation at 5000xg for 15 minutes at 4°C. The supernatants were discarded, and the pellets were stored in 50ml falcon tubes (pellets from 1L of culture per tube) at -20°C, at least overnight.

2.3 Sonication

The *E. coli* cell pellets (see above) were thawed and vortexed in 20ml of NMR buffer without D_2O; 200µl of Triton was added to each tube. The tubes were kept on crushed ice during sonication (~30% power, 15 seconds on/45 seconds off, 15 times). The sonicated samples were centrifuged at 10,000rpm for 25 minutes at 4°C; the supernatant was stored at 4°C for further purification.

2.4 Concentrating

The cell lysates were concentrated using an Amicon concentrator (5kDa Amicon cell stirrer filter) to ~5ml. The lysate was carefully filtered using a sterile 0.45 µM syringe filter.

2.5 Protein Purification

Two different methods were utilized for protein purification. Both methods used sized exclusion chromatography followed by ion exchange chromatography.

The original protocol utilized a Pharmacia Biotech P-50 protein pump to pump buffers. This pump was paired with a GE xk (16/100) column packed with Sephacryl S-100HR for size exclusion, and a GE xk (16/60) column packed with SP Sepharose fast flow resin for ion exchange chromatography. In an effort to increase efficiency of the purification method, we developed an alternate purification protocol using a BIO-RAD NGC Chromatography system, a HiPrep 16/60 Sephacryl S-300 HR column (gifted from Dr. Hans Schmitthenner, RIT School of

Chemistry and Materials Science) for size exclusion, and an SP FF 16/10 column for cation exchange chromatography.

2.5.1 Size exclusion Chromatography using Pharmacia Biotech P-50 Protein Pump

The filtered cell lysate was loaded onto the Sephacryl S-100HR XK (16/100) column at a flow rate of 2.5 ml/min. An isocratic elution using 50mM sodium acetate buffer pH 6.8 as the mobile phase and the parameters from (Table 2) were used to elute the protein. Fractions were collected in 30 ml borosilicate test tubes using the automated fraction collector.

Table 2 Program Method for Size Exclusion using GE XK 16/100 Columns

Break Point	**Volume (ml)**	**Flow Rate ml/min^{-1}**	**Fraction Volume (ml)**
1	0	2.5	0
2	600	2.5	0
3	601	2.5	22
4	2400	2.5	22

After each run, the column was washed using 0.5M NaOH, followed by re-equilibration of the column with 50mM sodium phosphate buffer, pH 6.8, at a rate of 2.5ml/min.

2.5.2 Size exclusion Chromatography using BIO-RAD NGC Chromatography Pump

The filtered cell lysate was loaded onto the HiPrep 16/60 Sephacryl S-300 HR column at a volume of 2.5ml per run using a Luer Lock syringe at a flow rate of 1.0 ml/min. An isocratic

elution using 50mM sodium acetate buffer pH 6.8 as the mobile phase and the parameters from (Table 3) were used. Fractions were collected in 8ml borosilicate test tubes using the automated fraction collector.

Table 3 Program Method for Size Exclusion Using FPLC

Break Point	Volume (ml)	Flow Rate ml/min^{-1}	Fraction Volume (ml)
1	0	1	0
2	10	1	0
3	11	1	6
4	210	1	6

2.5.3 Cation Exchange Chromatography using Pharmacia Biotech P-50 Protein Pump

Fractions from the SEC column containing the γB-crystallin protein were pooled and loaded onto the GE xk (16/60) column packed with SP Sepharose fast flow resin at a flow rate of 10 ml/min. Unbound protein was eluted in 0.275 M sodium acetate buffer, pH 4.6 at 10 ml/min. Sodium acetate buffer with NaCl was used to elute the protein, as described in (Table 4); 22 ml fractions were collected in 30 ml borosilicate test tubes.

Table 4 Programed Method for Cation Exchange Using XK 16/60 Column

Break Point	Volume (ml)	Salt Buffer Concentration	Flow Rate (ml/min)	Fraction Volume

		(% v/v)		(ml)
1	0	0	10	0
2	1800	0	10	0
3	1801	15	10	0
4	2860	27	10	0
5	3073	29	10	22
6	3180	30	10	22
7	3680	30	10	22
8	4704	43	10	22
9	4849	45	10	22
10	4850	85	10	22
11	4937	85	10	22

2.5.4 Cation Exchange Chromatography using BIO-RAD NGC Chromatography Pump

Fractions from the SEC step containing the γB-crystallin protein were pooled and loaded onto the SP FF 16/10 at a flow rate of 2 ml/min. Unbound protein was eluted using 0.275 M sodium acetate buffer at pH 4.6 at 2 ml/min. Sodium acetate buffer with NaCl was used to elute the protein, as described in (Table 5); fractions were collected in 8 ml borosilicate test tubes using the automated fraction collector.

Table 5 Programed Method for Cation Exchange Using FPLC

Break Point	Volume (ml)	Salt Buffer Concentration (% v/v)	Flow Rate (ml/min)	Fraction Volume (ml)

1	0	0	2	0
2	82.5	0-50	2	6
3	112.5	50-100	2	6
4	212.50	100	2	6

2.6 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE gels were used to determine which fractions contained the purified γB-crystallin protein. A 10% recipe [resolving gel: 3.27 ml of 30% acrylamide/bis-acrylamide, 3.33 ml Tris/SDS (Tris/SDS solution: 182 g Tris base, 1.5 g SDS, pH 8.0), 1.38 ml nanopure water, 2.12 ml 50% glycerol, 10% ammonium persulfate (APS), 10 µl tetramethylethylenediamine (TEMED); stacking: 405 µl 30% acrylamide/bisacrylamide, 775 µl Tris/SDS, 1.95 ml nanopure water, 20 µl 10% APS, and 5 µl TEMED] provided adequate separation at the lower molecular weight range (10-35kDa). BioRad gel systems were used to cast the gels, and 1x running buffer (10x running buffer: 30.0 g tris base, 144.0 g Glycine, 10.0 g SDS in 1 L H_2O, diluted 1/10 in nanopure H_2O) was inside and outside the gel chamber.

The samples were prepared by mixing protein and 2x sample buffer (4 ml of 10% SDS, 2 ml Glycerol, 1.2 ml of 1 M Tris pH adjusted to 6.8, 2.8 ml of H2O, 0.001-0.002 g bromophenol blue) at a 1:1 ratio; samples were boiled for 15 minutes.

Either 10 or 15 well gels were used to separate the protein samples; all samples were loaded into the gel at a volume of 14 µl. Eight microliters of Kaleidoscope protein ladder (BioRad and ThermoScientific) was also loaded as a molecular weight reference. The gels were run at 120v for 20 minutes, then at 150V for an additional 45minutes for adequate separation.

The gels were rocked in ~50ml InstaBlue (Fisher) for 1 hour to overnight to visualize the protein bands.

2.7 Buffer Exchange for Ion Exchange Chromatography

The size exclusion chromatography fractions that contained the γB-crystallin protein were combined into an Amicon stirred cell unit with a 5kDa filter. The total volume was brought down to ~4ml. The protein was buffer exchanged into 0.275M sodium acetate buffer, pH 4.6 by adding 4ml of the sodium acetate buffer to the solution, then concentrating back to 4ml. This was repeated 8 times to accomplish complete buffer exchange. The sample was removed from the amicon and syringe filtered using a 0.45 μM syringe filter.

2.8 Concentration via Centrifugation

Ion Exchange fractions containing γB-crystallin were placed into an Amicon Stirred Cell unit with a 5kDa filter. The total volume was brought down to ~4ml. The protein was exchanged into NMR buffer with D_2O and DTT by performing a similar dilution method as described in the previous section. The sample was removed from the Amicon stir unit and transferred into an Amicon Ultra-15 Centrifugal Filter Unit (Merck Millipore Ltd.). The protein was concentrated at 6,000*xg* until the sample reached a final volume of ~350 μl.

2.9 UV Spectroscopy

UV Spectroscopy was used to determine purity and concentration of the purified products. Samples were analyzed on a Shimadzu UV-VIS Recording Spectrophotometer (UV-2401PC).

Samples were analyzed using a scan rate of 240 nm min^{-1} in intervals of 0.1 nm for a wavelength range between 240–320 nm. NMR buffer without DTT was used to dilute the sample to avoid UV interference (5µl of the purified protein + 995µl of NMR buffer). The absorbance at 280nm was measured to determine the concentration (c) of the protein using the Beer-Lampert equation ($A=\varepsilon lc$) where A is the absorbance at 280nm, ε is the molar extinction coefficient (2.18 L / mol cm for γB-crystallin), l is the path length of the quartz cuvette (1cm).

2.10 NMR Spectroscopy

2.10.1 T_1/T_2 HSQC

All T_1 and T_2 relaxation NMR data were collected on the University of Rochester's 600MHz Varian spectrometer. We determined T_1 and T_2 values using the standard ^{15}N HSQC experiment from the Varian Protein Pack Library. To estimate these values for the global protein, we performed the experiment without allowing evolution on the nitrogen ($n_i=1$). We analyzed the 1D spectrum (projection along the proton axis) and integrated under the peaks in the 8-10ppm region. The following parameters were used in the T_1 experiment: nt=64, ss=8, np=1024, and relaxT=0.01, 0.05, 0.09, 0.15, 0.25, 0.5, 0.75, 1.0, 1.5, 2.3, 3.0, 4.0, and 5.0. The same parameters were used in the T_2 experiment except the relaxT values were varied to relaxT=0.01, 0.03, 0.05, 0.07, 0.09, 0.11, 0.15, 0.19, and 0.21. For the 2D HSQC experiment, the following parameters were utilized, $n_i=64$, nt=16, np=1024, ss=8. 12 different HSQC's were collected with the following relaxT values: 0.01, 0.05, 0.09, 0.15, 0.25, 0.5, 0.75, 1.0, 1.5, 2.3, 3.2, and 4.5.

After the raw T_1 and T_2 relaxation data were collected, the Varian software (VNMR 6.1.C) fit the T_1 data to a single exponential fit and determined both T_1 and T_2 time constants. Using the

T_1 and T_2 time constants, the τ_c values were calculated from a macro software package installed on the instrument using the approximation (Eq. 7.4.2), which was originally derived by Farrow et al. (1995 J Biomol. NMR).[44]

2.10.2 NMR Sample Preparations and Dilutions

Prior to running the NMR experiments, 325µl of protein were placed into a 5mm D_2O/Bruker Shigemi NMR Tube (Wilmad Lab Glass, USA). The sample was sealed using parafilm to avoid contamination.

Dilutions were performed on the NMR samples in order to determine relaxation times over a range of protein concentrations. The highest concentration was always measured first. Once the data was analyzed for a specific concentration, a 1:1 dilution was performed in NMR buffer (with D_2O and DTT) in a microfuge tube. The sample was gently mixed by vortexing, before being placed back into an NMR tube. This process was repeated multiple times to study a range of protein concentrations.

2.11 Calibration Standard

Prior to analyzing the DOSY data, we performed a calibration of the gradient using a standard sample: 4.64 mM solution of sucrose in D_2O (7.94 mg of Sucrose in 5ml of 99% D_2O).

2.12 Calculations

All calculations and mathematical analyses were performed using Wolfram Mathematica 11 Student Edition (Version 11.2.0.0).

3 Results

3.1 SDS PAGE Gel

SDS PAGE gels were employed after each column run to identify and track the γB-crystallin protein throughout the protein expression and purification processes. The gel below (Fig. 4) contains the fractions collected from the size exclusion column. γB-crystallin has a molecular weight of approximately 21kDa, as seen in the thick band between 17 and 26kDa. The presence of the thick band suggests either degradation of the protein, or coelution of another protein with similar size. Most of the heavier proteins eluted in the earlier fractions, suggesting the size exclusion technique worked as intended.

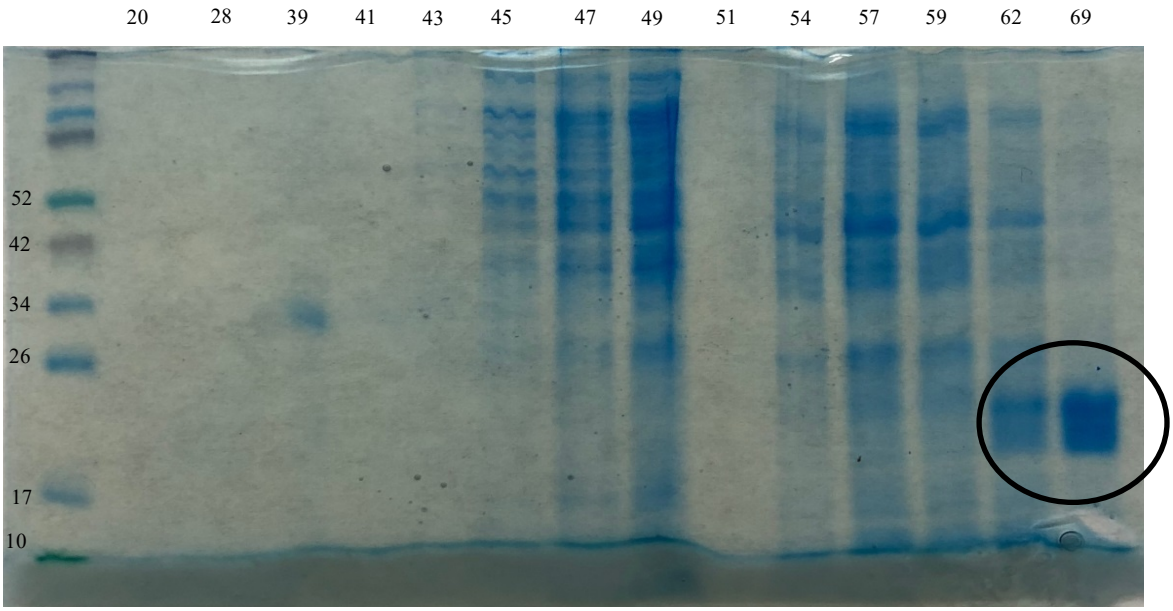


Figure 4 SDS-PAGE Gel After Size Exclusion Chromatography

An SDS-PAGE confirming successful expression and purification of the 21kDa γB-crystallin (circled in black), located in fraction 62-69.

Further purification was performed using ion exchange chromatography to remove proteins of similar size, but different charge:

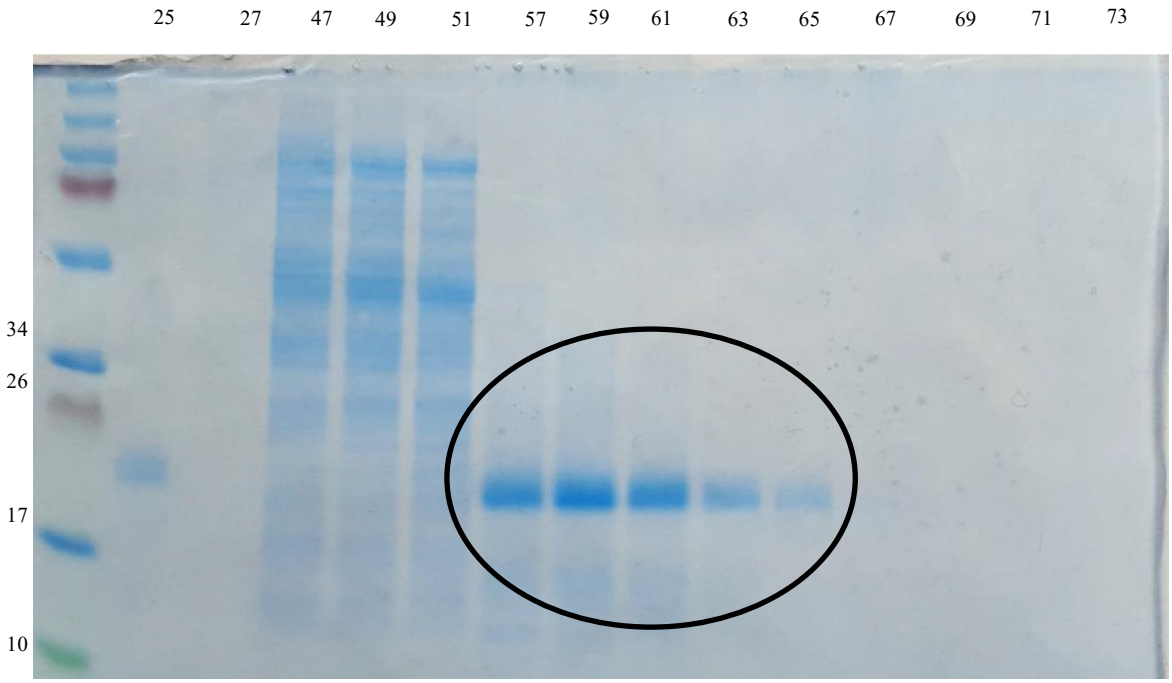


Figure 5 SDS-PAGE Gel After Ion Exchange Chromatography

An SDS-PAGE confirming successful expression and purification of the 21kDa γB-crystallin (circled in black). Fraction 25: contaminant, fraction 27: empty, fraction 47-51: waste, fraction 57-65: purified γB-crystallin eluting over multiple fractions, fraction 67-73: empty.

3.2 UV Spectroscopy

UV spectroscopy was used to measure the protein concentration of our samples, as well as to confirm purity. Using Beer's Law, $A=\varepsilon bc$, with a pathlength (b) of 1 cm and a molar extinction coefficient (ε) of 2.18 L mol^{-1} cm^{-1}, the concentration of each protein sample was determined based on its absorbance at 280nm. As seen below (Fig. 6), γB-crystallin has a unique characteristic absorption spectrum that we used to help verify protein purity.

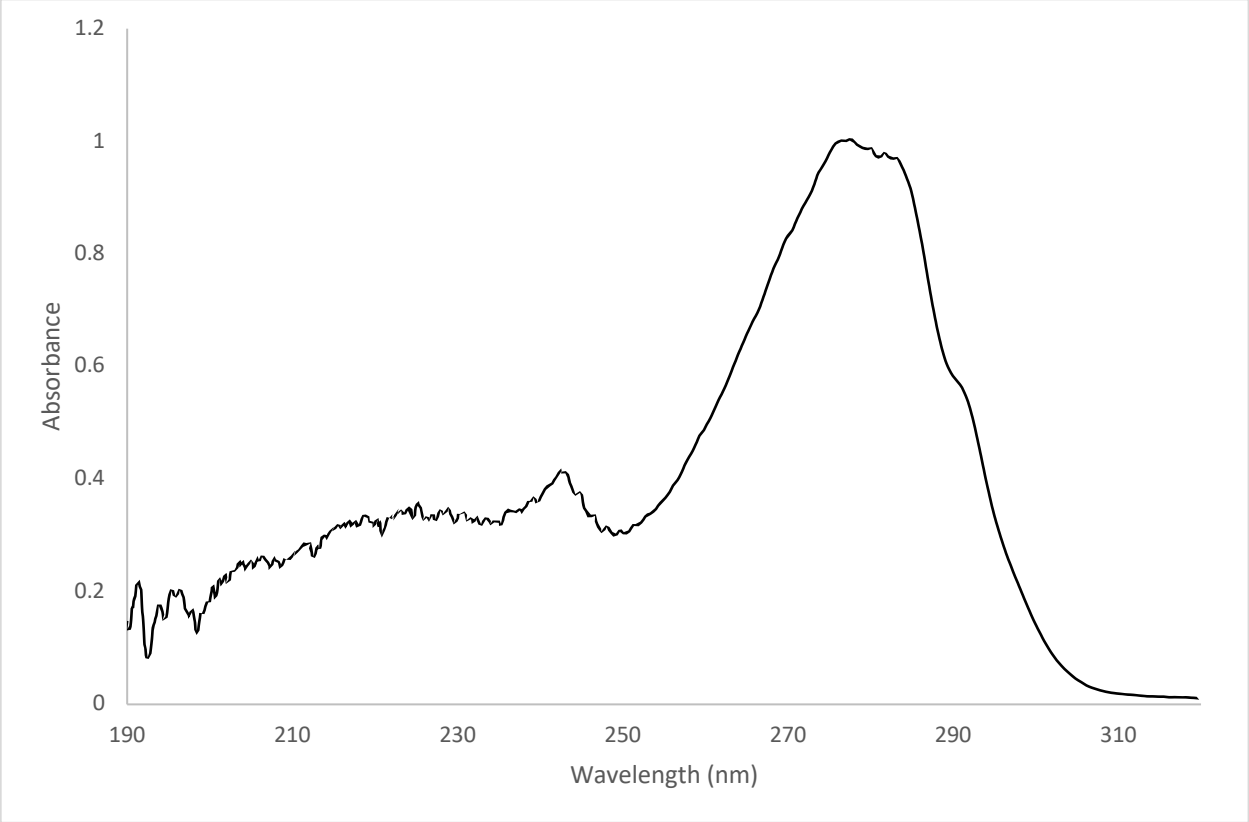


Figure 6 UV-Absorbance Spectrum of γB-crystallin

UV-absorbance spectrum of purified γB-crystallin protein. The shoulder at ~290nm, along with broad peak at 280nm is characteristic of the γB-crystallin.

3.3 T_1/T_2 NMR Results

3.3.1 τ_c Times, Single Exponential Analysis

The Rotational Correlation Values, τ_c, were measured using T_1 and T_2 relaxation NMR on the University of Rochester's 600MHz Varian spectrometer. First, we plotted the τ_c data calculated from the macro package discussed in the Section (2.10.1). The macro package provided us an analysis that contained a τ_c value, as well as an error analysis (shown with error bars) and assumes a single exponential decay for the T_1 data. Two different rounds of ^{15}N labeled γB-crystallin protein was grown and purified resulting in starting concentrations of 33 mg/ml and 28 mg/ml. τ_c times were measured at different temperature and subsequent dilutions were performed. The results are as follows:

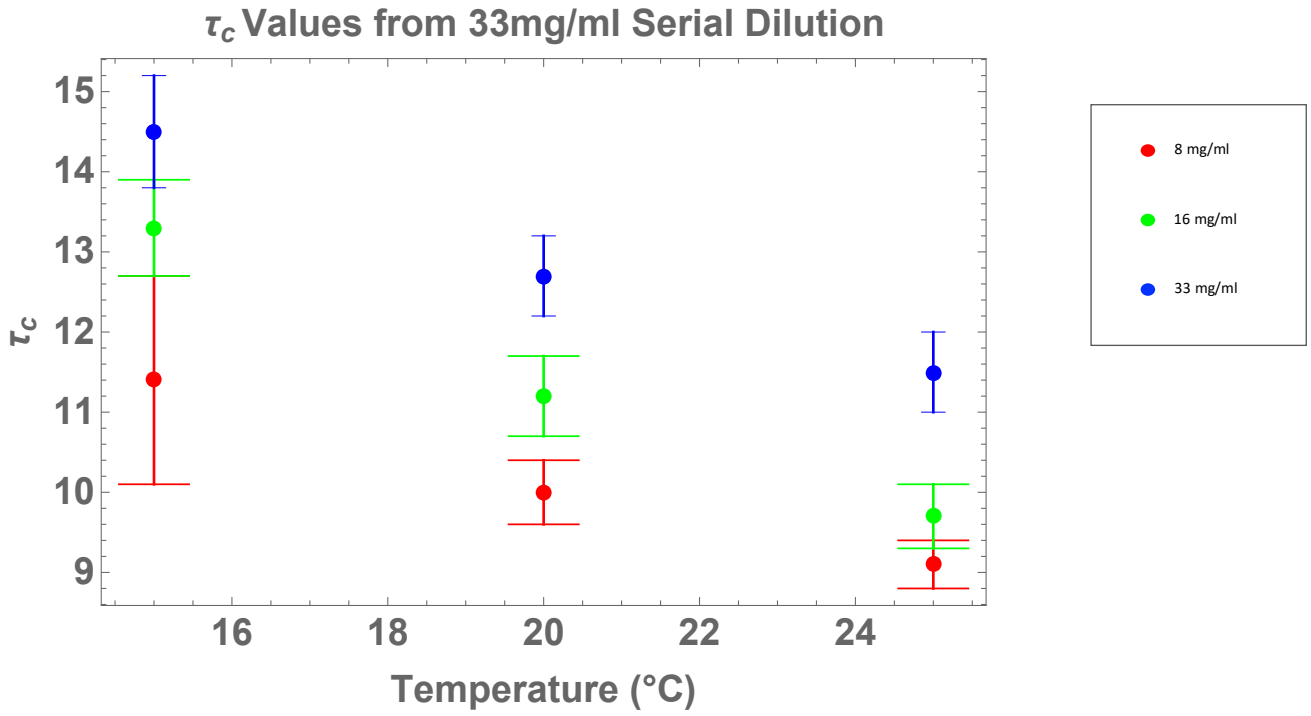


Figure 7 Varian τ_c Values from 33 mg/ml Serial Dilution

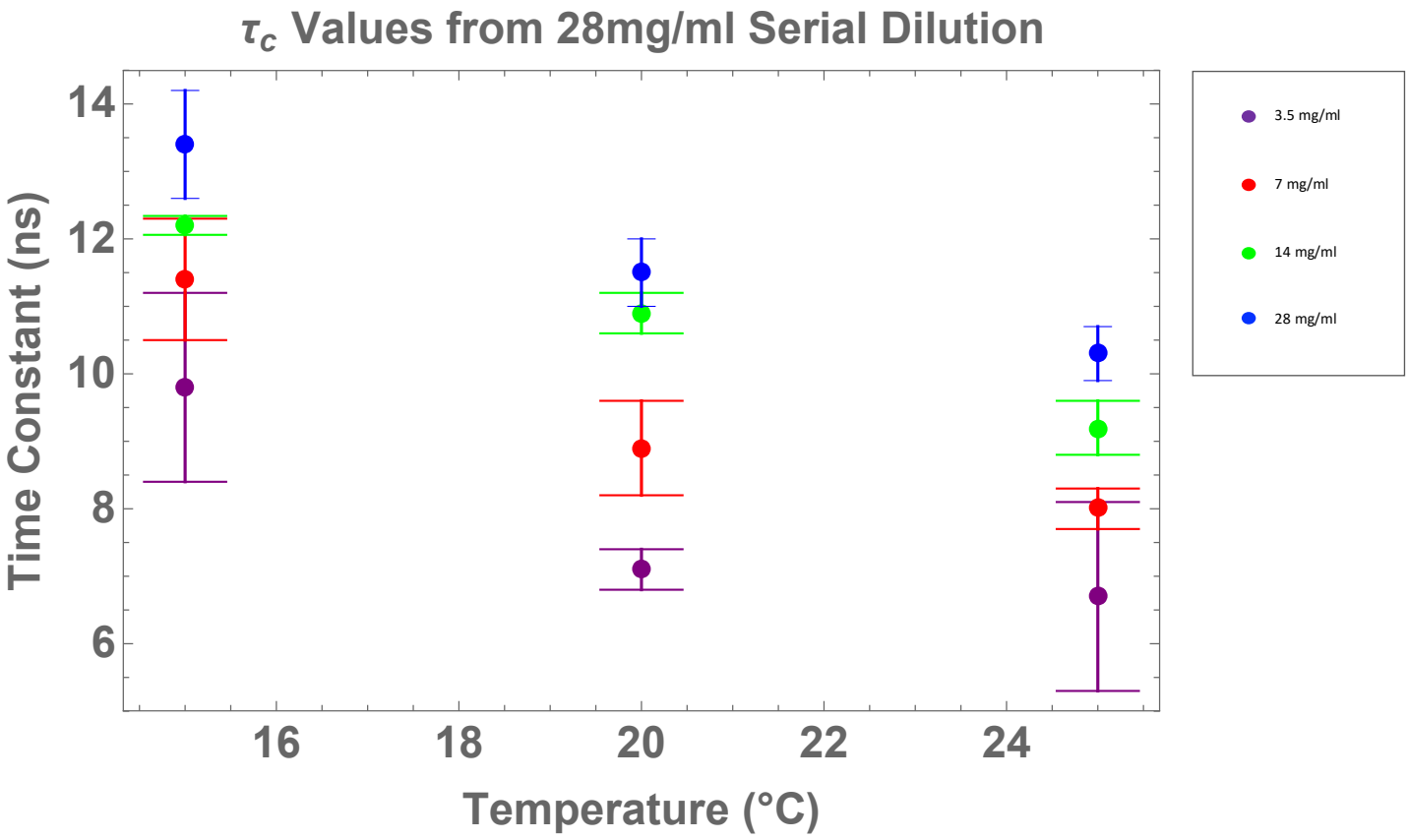


Figure 8 Varian τ_c Values from 28 mg/ml Serial Dilution

3.3.2 (T_1) Double Exponential Analysis

Although the macro package installed on the Varian assumes a single exponential decay to calculate T_1, we noticed that a single exponential decay did not fit the data as well as expected. Upon further inspection, we discovered that a double exponential decay fit the T_1 data better. This was an important discovery, because it altered the effective T_1 time constant values, which therefore impacted the τ_c values.

By analyzing the raw data using a Mathematica script, we fit our decays using double exponentials, which resulted in two unique T_1 time constants. This Mathematica script utilized the function "NonlinearModelFit" for the initial fitting of the data. To calculate the standard error, "ParameterErrors", a separate function that is built into the "NonlinearModelFit" function was used. This function assumes a normal distribution of errors. To determine the scale of variance, another default function, "VarianceEstimatorFunction" was applied, which uses equal weighting of each data point. The variance scale was given by the sum of the weighted, squared residuals for the fit, divided by the difference between the number of data points, and the number of parameters used for the fit.

These two time constants were very different from each other, with time constant 1 being close to 1-1.5s (well within the expected range), and the second time constant ranging from 0.05-0.2s, much faster than then typical T_1 relaxation times. There was a clear trend in the time constant 1 data, as seen in the example below (Fig. 9). Time constant 2 did not show as clear of a trend and remained stable regardless of temperature and concentration.

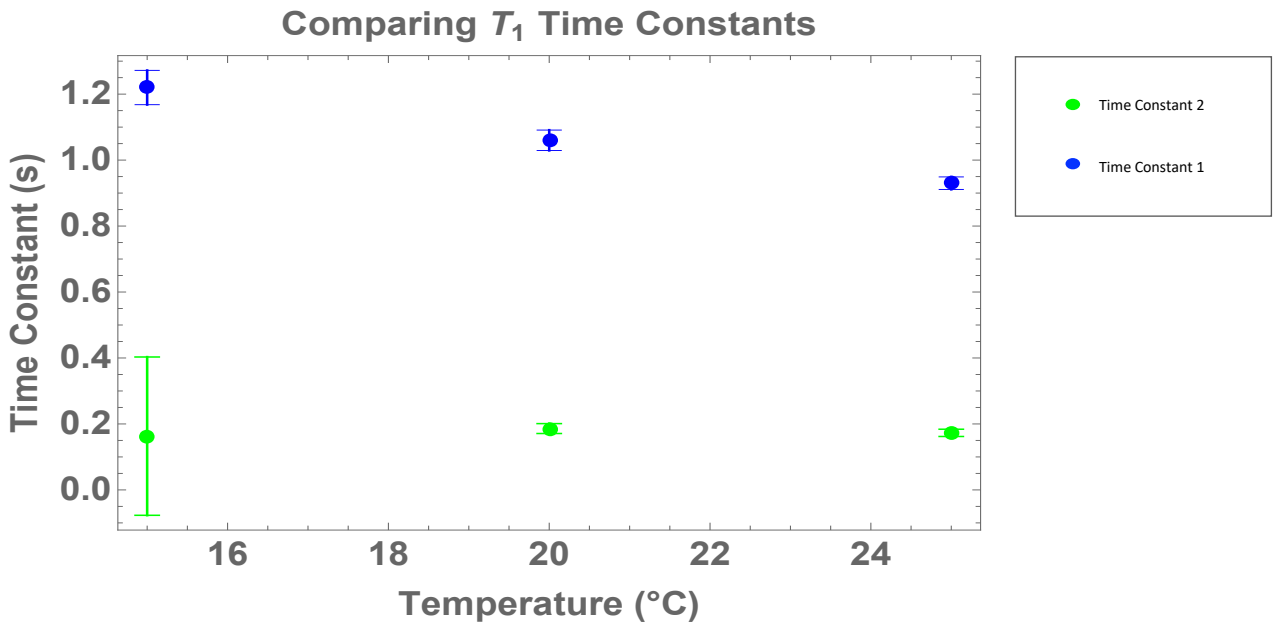


Figure 9 Time Constant 1 and 2 from T_1 Experiment

Fitting a double exponential decay for the T_1 relaxation data yielded two separate time constants, one of which was much faster than previously measured, and well short of the average T_1 times for biological systems.

3.3.3 τ_c Times Double Exponential Analysis

Fitting the T_1 data from above using double exponential decays yielded two separate time constants for T_1 relaxation. Using the longer of the two time constants as the value for T_1, the approximate τ_c values were calculated (Fig. 10) using (Eq. 7.4.2). The data point for the 33 mg/ml, 15°C sample was omitted, because the T_2 NMR data was not saved correctly, and the τ_c value could not be calculated using (Eq. 7.4.2) without the T_2 time constant. Error bars were calculated using a propagation of error from the T_1 and T_2.[45]

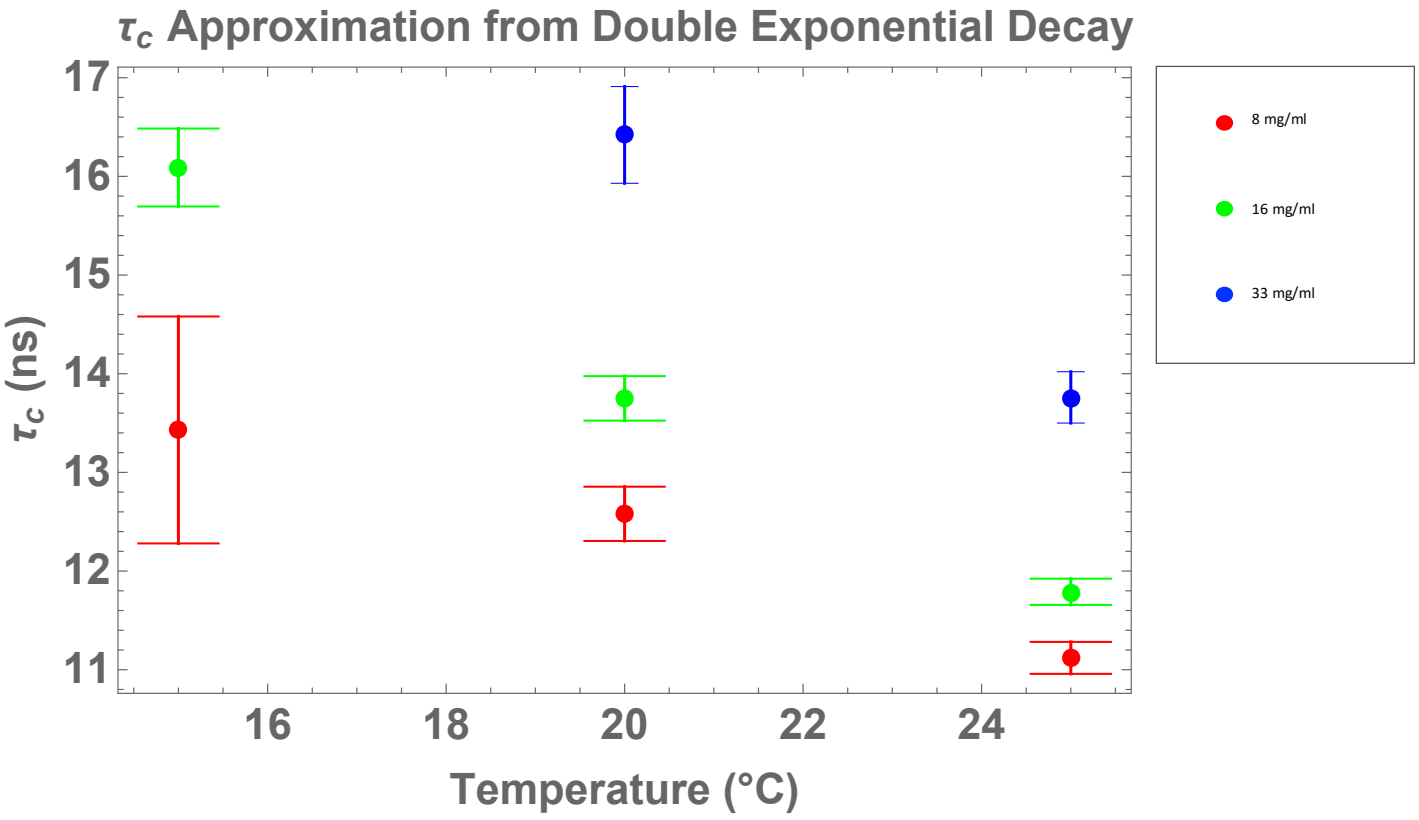


Figure 10 Rotational Correlation Times from Double Exponential Decay

Rotational Correlation Times (τ_c) calculated from the T_1/T_2 NMR values. Measurements were recorded at various temperatures and concentrations, and the analysis was performed using the longer time constant calculated using a double exponential decay.

3.3.4 Comparison of Varian Approximation with Full τ_c Formula

To confirm the accuracy of the rotational diffusion measurements calculated on the Varian NMR, the full τ_c equation was derived by Dr. George Thurston from the relaxation constants T_1 and T_2. Both the approximated τ_c values (Eq. 7.4.2) and the fully worked out τ_c values (ratio of T_1 and T_2 spectral density equations) were compared (Fig. 11). If the approximated equation was accurate, then using the same T_1 and T_2 values should have produced highly similar or identical τ_c values over a wide range. The values were overlaid in the plot below (blue and green lines). As seen in the plot, the τ_c approximation formula that the Varian software uses (Eq. 7.4.2) produces values well within an acceptable range for τ_c.

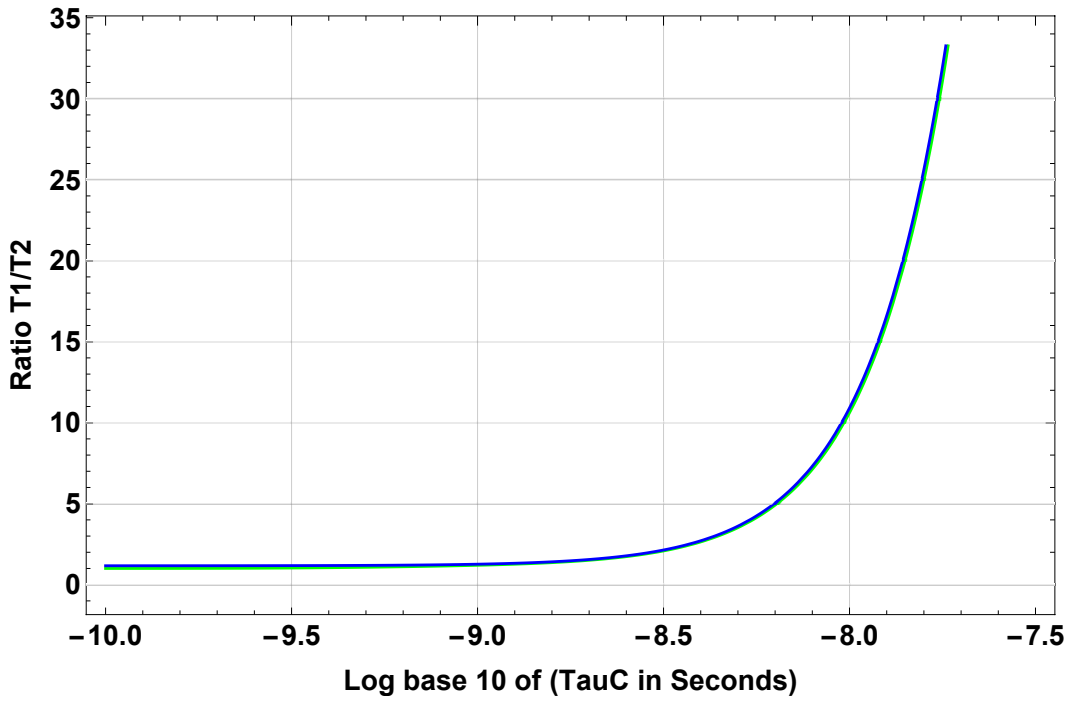


Figure 11 Comparison of Varian Approximation with Full τ_c Formula

Log base 10 of τ_c was compared for the approximated τ_c Varian formula (Blue) to the fully derived τ_c equation (Green).

Additionally, to mathematically determine how close the approximation is to the fully derived τ_c value, the ratio of the two calculated values were plotted below (Fig. 12). The plot confirms the estimated τ_c value is very close to the fully derived τ_c. formula.

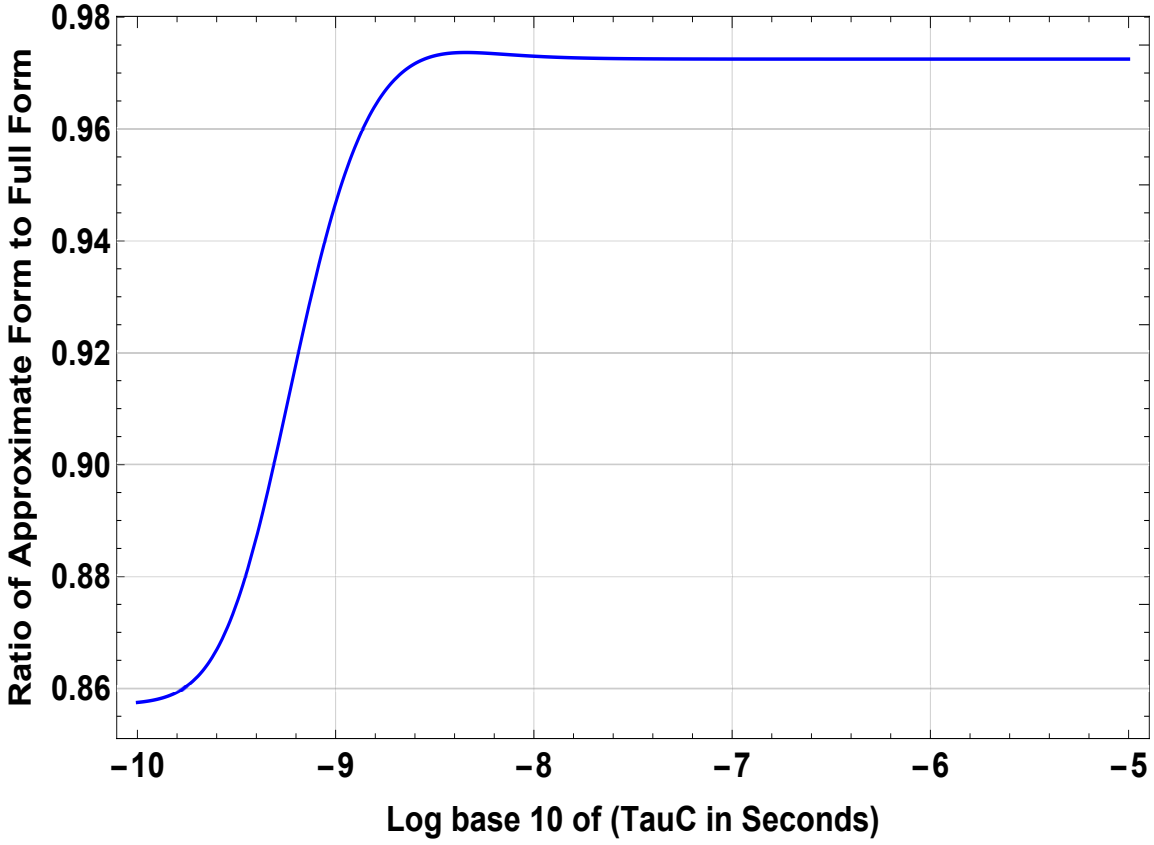


Figure 12 Ratio of Approximated τ_c Values to Actual Values

3.3.5 (T_1) Time Constants with a Single Exponential Decay

To better understand the τ_c data, the T_1 time constants, were graphed vs. temperature at the different concentrations using the most recent NMR data. Using a single exponential decay analysis, the time constants followed a clear trend, in which the T_1 time constants increased as temperature decreased and increased as concentration increased (Fig. 13). The range of time constants for the T_1 analysis ranged from 0.69 up to ~1sec, which are typical T_1 relaxation times for proteins.

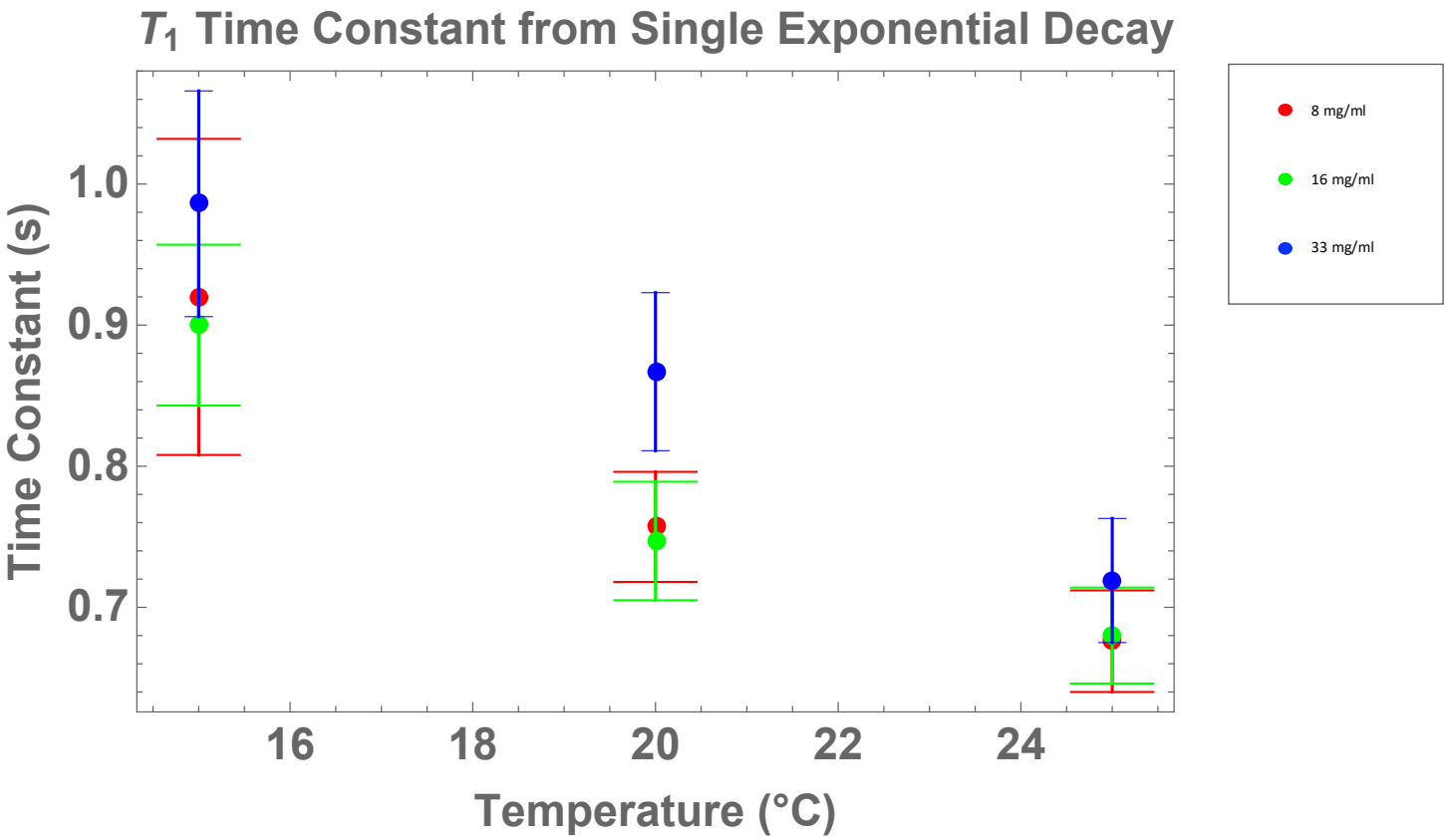


Figure 13 Time Constant 1 from T_1 NMR

Comparison of time constants from single exponential decay analysis at various temperatures and concentrations.

3.3.6 (T_1) Time Constant 1 from Double Exponential Decay Analysis

The longer of the time constants were plotted to determine the effects of temperature and concentration. A clear and expected trend was measured for this time constant across all the concentrations and temperatures (Fig. 14). As temperature was decreased, the time constant increased, and as concentration decreased, so did the time constant. The time range for the first-time constant falls between 0.9-1.5sec, a similar range as the T_1 calculated using a single exponential decay.

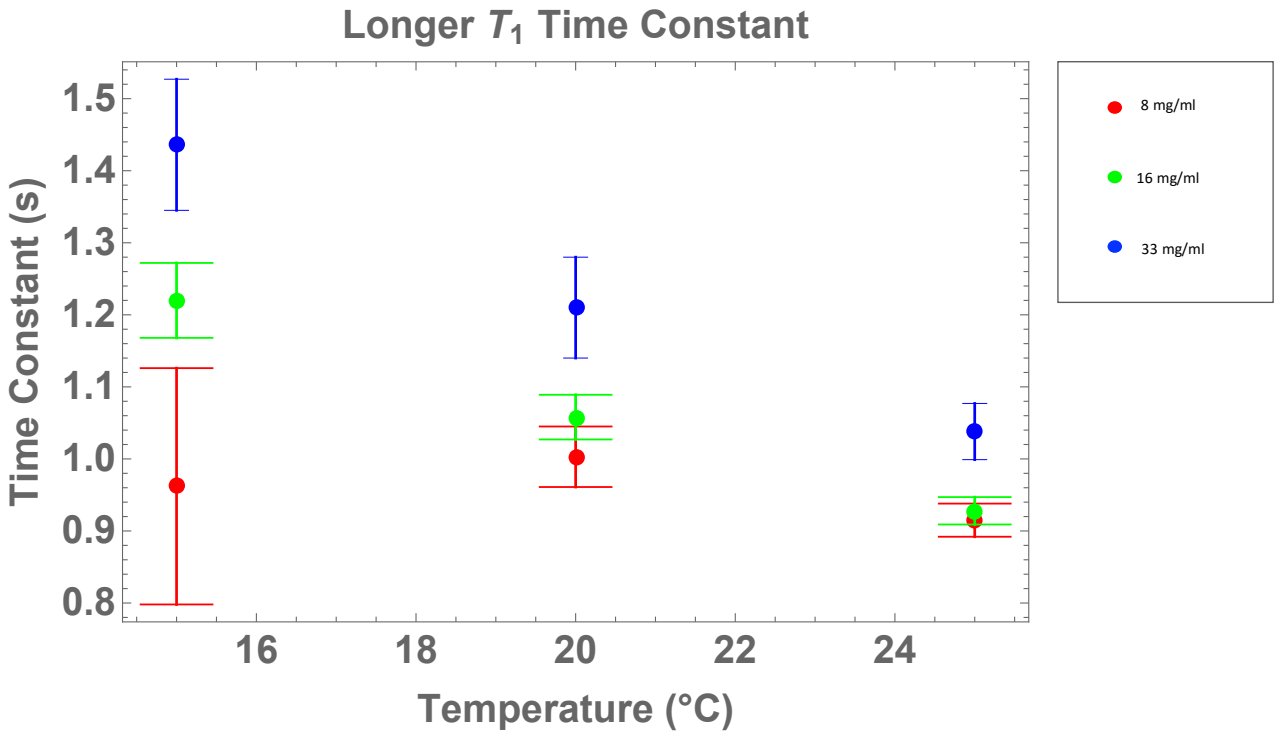


Figure 14 Time Constant 1 from T_1 Double Exponential Decay

Comparison of time constant 1, the longer of the two-time constants resulting from a double exponential decay, plotted at various temperatures and concentrations.

3.3.7 (T_1) Time Constant 2 with a Double Exponential Decay

The shorter of the time constants were also plotted to determine how temperature and concentration affected the relaxation time (Fig. 15). The second time constant was much faster, with times ranging from 0.05 to 0.5sec, which are very short values for T_1 time constants in biological systems. Trends within the second time constant were much less apparent compared to the first time constant, although the data suggest there may be a correlation between concentration and time constant.

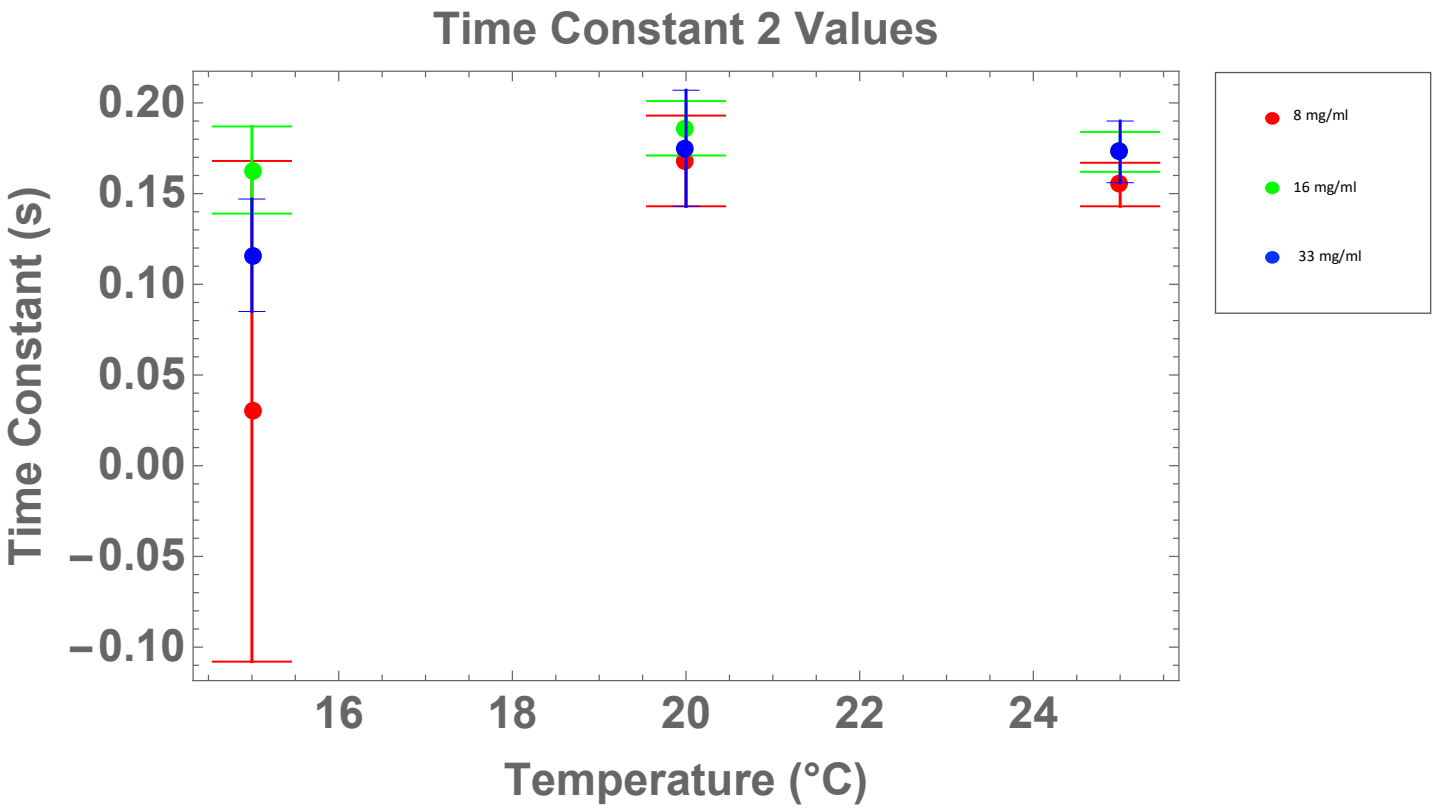


Figure 15 Time Constant 2 from T_1 Double Exponential Decay

Comparison of time constant 2, the shorter of the time constants resulting from a double exponential decay, plotted at various temperatures and concentrations.

3.3.8 T_2 Time Constant

The T_2 data were also analyzed to calculate the T_2 time constants. The results of this analysis are graphed below to show the relationship between temperature, concentration, and T_2 time constants (Fig. 16). The data below show that higher concentrations will have shorter T_2 time constants. The 33 mg/ml, 15°C sample data were not properly saved, so that data point was omitted from analysis.

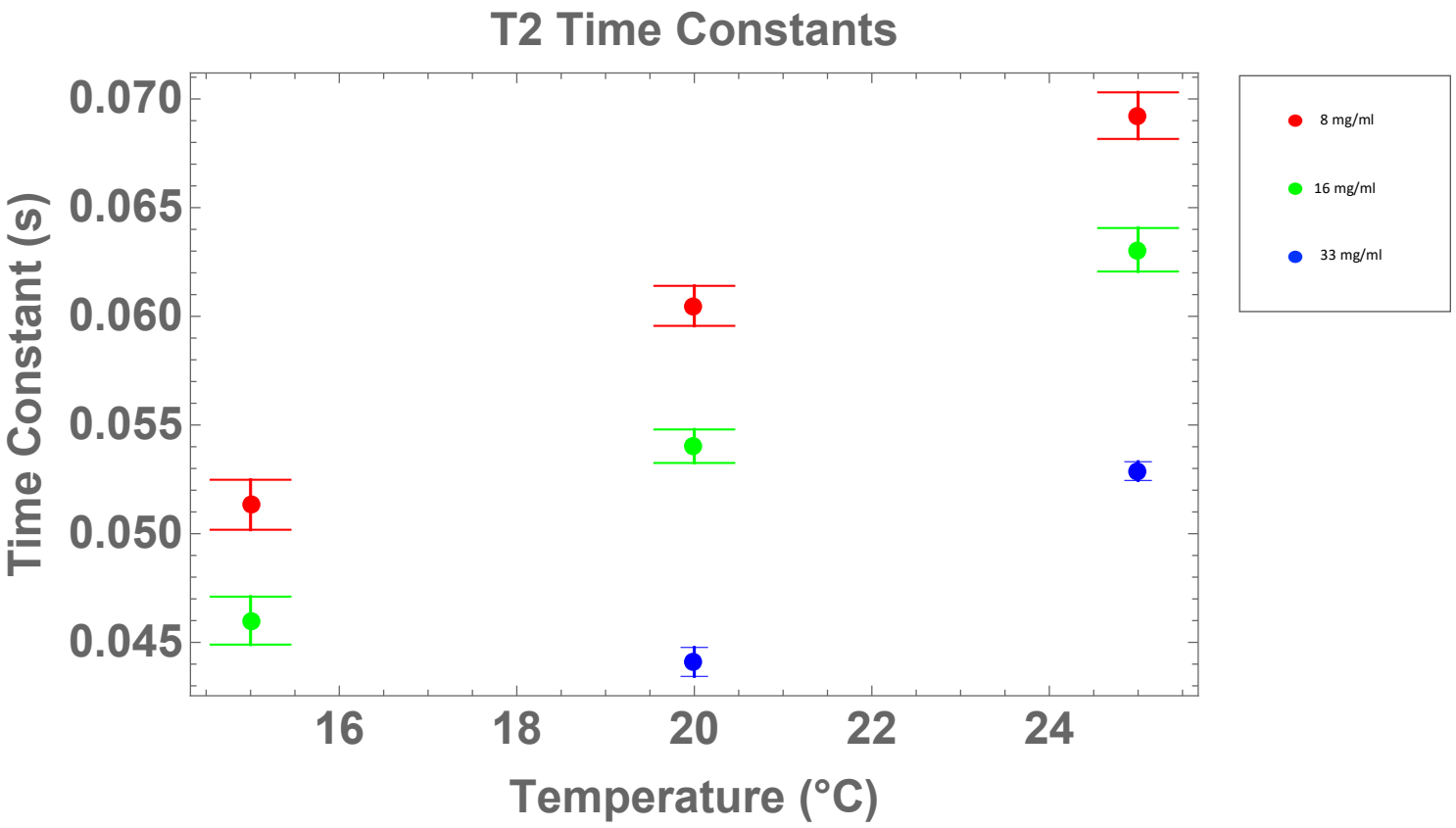


Figure 16 T_2 Time Constants

3.3.9 Hydrodynamic Radii

In an effort to glean insight into the aggregation of bovine γB-crystallin, the approximate hydrodynamic Radii and corresponding error bars were estimated using the T_1 and T_2 data and formula 7.4.1 (Fig. 17). The calculated values suggest that hydrodynamic radius is affected more by protein concentration than by temperature, and the average hydrodynamic radius of the bovine γB-crystallin falls around 2nm.

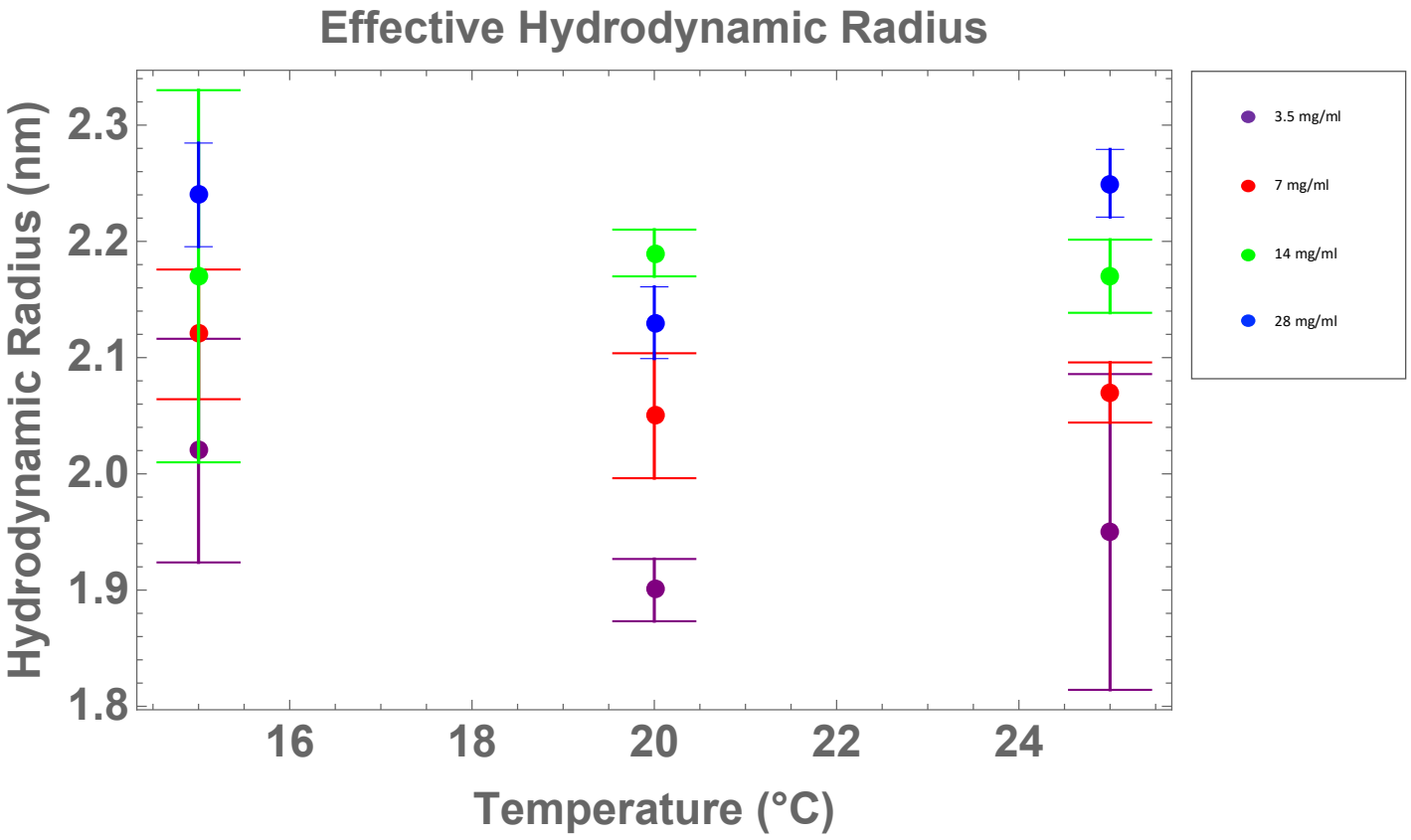


Figure 17 Hydrodynamic Radius Calculated from τ_c times.

3.3.10 Durbin Watson Single Vs. Double Exponential Analysis 33 mg/ml

To determine whether the single or double exponential decay constants fit the data better, we implemented the Durbin Watson statistical test. This test is used as a measure of autocorrelation within a regression analysis. Using the raw numerical data, a Mathematica script was written to measure the Durbin Watson Statistic. This analysis was performed on data from multiple NMR samples. (Fig. 18) is an example of the results obtained from the Mathematica script. The top portion represents a single exponential decay analysis, while the bottom portion represents the data fit with a double exponential. The double exponential decay is a seemingly better fit. The sinusoidal nature of the residuals in the single decay data indicates the presence of autocorrelation, while the randomized residuals in the double decay data suggest a lack of autocorrelation. The Durbin-Watson values expressed next to the plotted residuals support this claim, as well (Fig. 18). Additionally, the R^2 values were calculated. These values, also known as the coefficient of determination, tell us the proportion of the variance found in the dependent variable that results from the independent variable. The R^2 value for the double exponential analysis (0.999) also suggests a better fit than the single exponential analysis (0.995). As described above, the double exponential fitting produces two time constants for T_1 relaxation, potentially indicating two separate populations, which are "relaxing" at different rates. The single exponential decay yields one time constant, which is possibly the weighted average of the two time constants from the double exponential analysis.

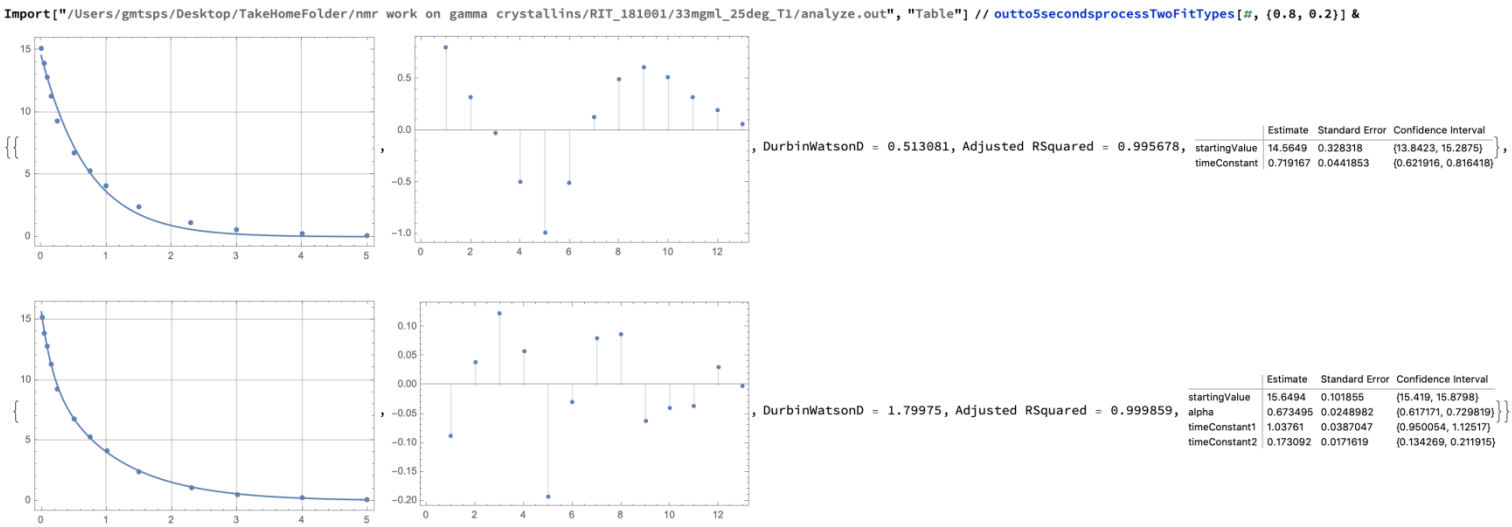


Figure 18 Mathematica Analysis of Single Vs. Double Exponential Decay of 33 mg/ml at 25°C

The top set of data utilizes a single exponential decay fitting, while the bottom set uses a double exponential fitting. From left to right: Relaxation times with exponential decay curve fitted, plotted residual points, calculated Durbin-Watson values, R^2 values, and time constants with error analysis.

Table 6 R^2 and Durbin-Watson Values for Single and Double Exponential Decay

Concentration (mg/ml)	Temperature (°C)	Single Exponential (d)	Double Exponential (d)	R^2 (Single Exponential)	R^2 (Double Exponential)
33	25	0.513	1.799	0.995	0.999
33	20	0.512	2.553	0.995	0.999
33	15	0.458	2.519	0.992	0.999
16	25	0.514	3.13	0.997	0.999
16	20	0.503	2.289	0.996	0.999
16	15	0.466	1.854	0.995	0.999
8	25	0.525	1.927	0.996	0.999
8	20	0.567	1.948	0.996	0.999
8	15	3.183	3.256	0.983	0.981

3.4 Akaike Information Criterion Results

The Akaike Information Criterion was applied to all of the available T_1 NMR data. The AIC values were calculated for both a single and double exponential model for the T_1 delays. Once calculated, the values for each model were used to calculate the *relative likelihood (i)* of a particular model in comparison to the other. For all temperatures and concentrations, the relative likelihood favored the double exponential decay, with the exception of the 8 mg/ml sample measured at 15°C, which favored the single exponential decay.

Table 7 Akaike Information Criterion Results

Concentration	**Temperature**	**Single Exponential Akaike**	**Double Exponential Akaike**	***Relative Likelihood of Single Exponential Fitting Better than Double Exponential (i)***
33 mg/ml	25	27.887	-8.695	1.13831×10^{-8}
33 mg/ml	20	27.887	4.495	8.82686×10^{-7}
33 mg/ml	15	-181.842	-204.173	1.41552×10^{-5}
16 mg/ml	25	24.763	-24.360	2.15273×10^{-11}
16 mg/ml	20	26.284	-14.698	1.26123×10^{-9}
16 mg/ml	16	25.7154	-2.108	9.08182×10^{-7}
8 mg/ml	25	38.9723	4.292	2.9471×10^{-8}
8 mg/ml	20	40.8131	12.592	7.44655×10^{-7}
***8 mg/ml**	15	58.0346	68.034	$1/6.73879 \times 10^{-3***}$

Akaike Information Criterion results for both single and double exponential fits of the T_1 relaxations. The relative likelihoods of each model were also calculated. All relative likelihoods favored the double exponential analysis, *except* the 8 mg/ml at 15°C sample denoted by (***)

3.5 Translational Diffusion Coefficient Measurements

A 4.64mM sucrose standard was prepared in 10% D_2O and analyzed via DOSY NMR. This standard was used to calibrate the gradient in our 500 MHz spectrometer at RIT. Gradient calibration involves measuring the actual strength of the gradient in G/mm. The NMR probe is preprogramed to assume a gradient strength of 5.3500094 G/mm. A DOSY experiment was run at 25°C to determine the diffusion coefficient of the sucrose. Using the Topspin software, diffusion coefficients were measured. These values are found below, along with the published diffusion coefficients values:

Table 8 Experimental vs. Published Diffusion Coefficients for Sucrose

Experimental	4.05×10^{-6}
Published Values [46]	4.92×10^{-6}
Difference	8.7×10^{-7}

The raw data collected for the 4.64 mM sucrose standard was plotted as peak integrals versus the gradient amplitude. The data fits a Gaussian shape as expected, showing a decrease in peak integral as the gradient amplitude increases.

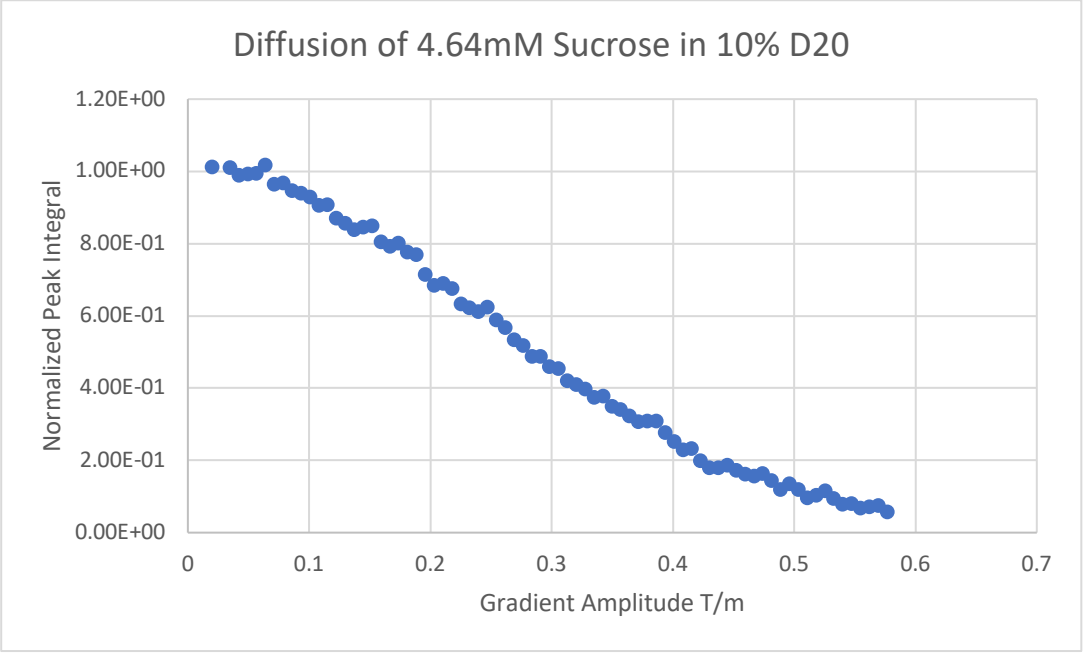


Figure 19 Peak Integral vs Gradient Amplitude

A Stejskal-Tanner plot was created from the raw data. This type of plot uses the Stejskal-Tanner Equation (Eq. 7.4.3) to linearize the data, producing a straight line with a slope that equals the diffusion coefficient, D. The value from the plot indicates a diffusion coefficient of 3.77 x 10^{-6}cm^2/s. Using the Linest function in excel, an analysis on the fit of the line was performed. This provided the error in the slope of the line, and since the slope of the Stejskal-Tanner plot is equal to the diffusion coefficient, it provides the error in our analysis of the Diffusion coefficient. Accordingly, using the Stejskal-Tanner plot in conjunction with a Linest function, we determined the diffusion coefficient to be 3.77 ± .026 x 10^{-6} cm^2/s. The Topspin software also produces a diffusion coefficient as part of the report. The value reported by the instrument was 4.05 x 10^{-6} cm^2/s, indicating that our method of determining diffusion coefficients from the slope of the Stejskal-Tanner plot was close, but not exact.

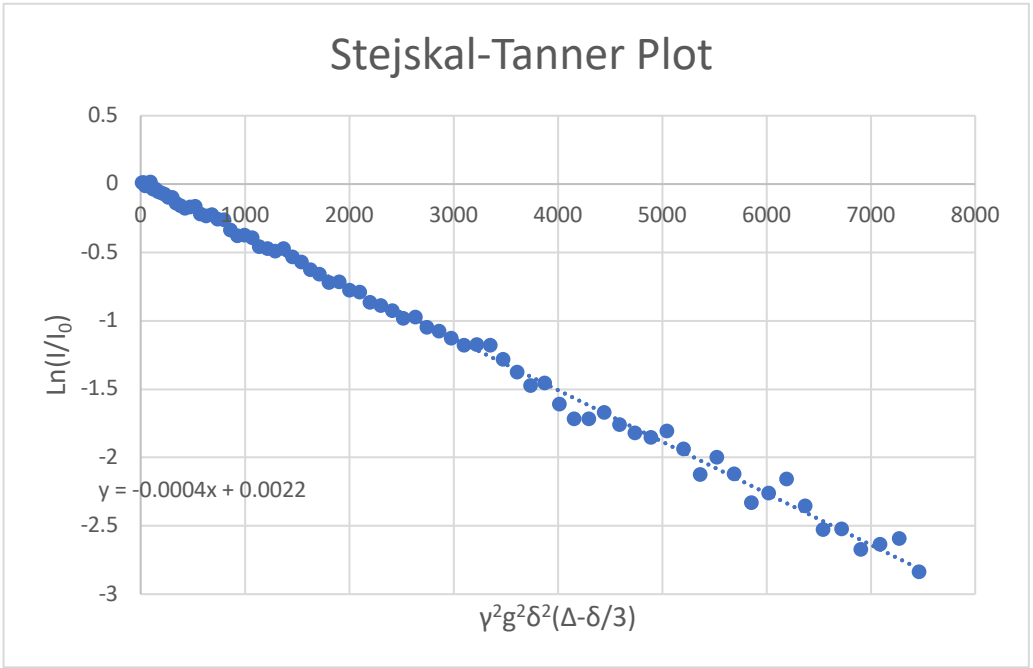


Figure 20 Stejskal-Tanner Plot for Sucrose Std.

Since gradient calibration only impacts how the data are analyzed, its effect can be accounted for on data that have already been collected. Using DOSY NMR, the diffusion coefficients of non-isotopically labeled bovine γB-crystallin were measured for several samples at different concentrations, ranging from 33 mg/ml to 4.45 mg/ml, at 25°C. At 4.45 mg/ml, the signal to noise (s/n) ratio was very low, so data from that sample were omitted in the analysis. The diffusion coefficients were measured for both the amide and aliphatic regions of the spectrum, then averaged to yield the overall diffusion coefficient for the protein. The following (Table 9) contains the experimental diffusion coefficients calculated from the DOSY experiments, which were collected ***before*** the gradient calibration. Due to issues with water suppression, we had to integrate the spectra upfield and downfield of the water peak, rather than a full integration. This resulted in diffusion coefficients for each section.

Table 9 Pre-Gradient Calibration Diffusion Coefficients for Different Concentrations at 25°C

Temperature	Concentration	Diffusion Coefficient for Downfield Region (cm^2/s) (4.8-10ppm)	Diffusion Coefficient for Upfield Region (cm^2/s) (0-4.7ppm)
25°C	33.00 mg/ml	6.10×10^{-8}	5.78×10^{-7}
25°C	24.47 mg/ml	6.25×10^{-8}	8.58×10^{-7}
25°C	12.23 mg/ml	4.55×10^{-7}	8.82×10^{-7}

To more easily visualize the trends in the translational diffusion data, the diffusion coefficients were plotted as a factor of protein concentration:

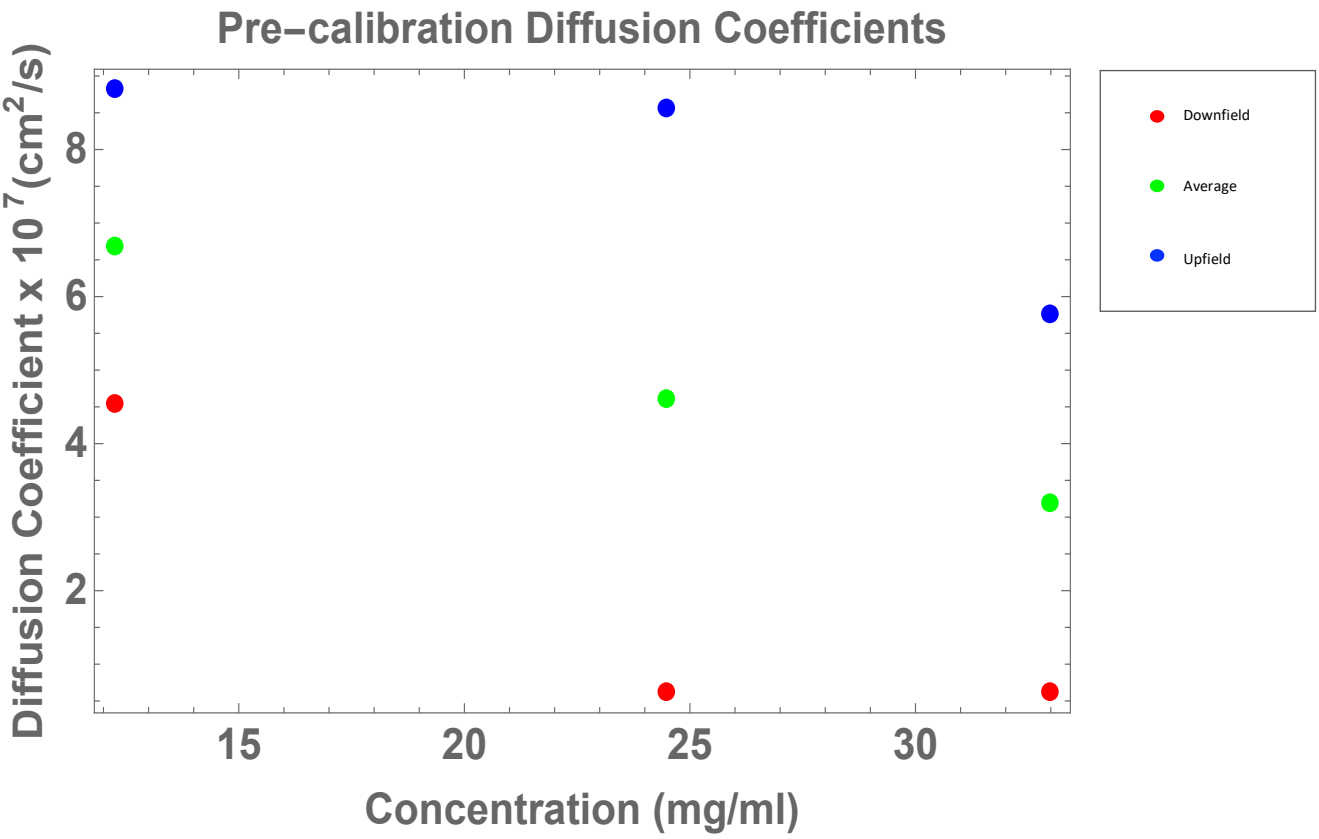


Figure 21 Pre-Calibration Diffusion Coefficients γB-crystallin

4 Discussion

4.1 Purification Process

Within the human eye, the γ crystallins naturally occur at high concentrations, close to 400 mg/ml. For this work, our highest concentration obtained was 41 mg/ml for ^{15}N-labeled γ-crystallin and (95 mg/ml) for unlabeled protein.

Initially, large scale production using the GE XK columns produced yields ranging from 30-40 mg/ml from 6 liters of growth in minimal media. In an effort to increase yields, we developed a new purification method using an FPLC instrument. However, this method did not result in better yields, and although it required less buffer, it was much more time consuming.

We used non-isotopically labeled protein for DOSY NMR to measure translational diffusion; non-isotopically labeled protein gives better yields and is less expensive to produce. Higher protein concentrations were desired for diffusion measurements for a variety of reasons. Firstly, to better understand how protein interactions within the eye are responsible for cataract formation, it is important to mimic the conditions of the eye as closely as possible. This means that the closer the concentration falls to 400 mg/ml, the more closely the analysis reflects how the proteins act in the eye lens. Secondly, higher concentrations of the proteins make NMR diffusion measurements easier. Higher concentrations create larger signals, meaning better signal to noise ratios. At the lower concentrations, more scans were necessary to accurately measure relaxation signals.

4.2 Rotational Correlation (τ_c)

Rotational correlation times can provide information on protein-protein interactions. T_1 and T_2 relaxation times measured over a range of temperatures and concentrations can be used to quantitatively calculate the rotational correlation time (τ_c). This value provides us valuable insight into how the protein is tumbling in solution. We focused our analysis on the backbone amide region of the NMR spectrum (8-10ppm region). Outside of this region, we can see amides that are located on the sidechains of amino acids such as Tryptophan, Asparagine, Glutamine, and Histidine. This leads to another potential interesting area of study. By monitoring the relaxation of specific sidechains of amino acids over time, it may be possible to gain insight as to whether certain residues are relaxing faster than others. In this case, perhaps there are specific regions of the protein that are highly disordered or that undergo motions separate from the rest of the protein; these motions would certainly have an impact on the overall τ_c value.

Because the rotational correlation value represents indicates the time it takes a globular protein to rotate a full radian, it can be used as an indicator of association rates of the protein in solution. When a single protein associates with another protein, its effective size and shape changes. The effective size of the protein will continue to increase as more molecules associate, therefore increasing the effective hydrodynamic radius and increasing the rotational correlation times. Because of this, it is expected that as more molecules associate, the rotational correlation time of this cluster of proteins will increase, indicating slower tumbling rates.

Two different physical conditions were studied for their effects on τ_c: concentration and temperature. Rotational correlation times were determined for the bovine γB-crystallin protein at concentrations between 33 mg/ml and 3.5 mg/ml and three different temperatures: 15°C, 20°C, and 25°C. Although these conditions do not accurately represent *in vivo* conditions within the

eye, this work provides the first rotational correlation measurements for the bovine γB-crystallin protein using T_1/T_2 NMR.

From our T_1/T_2 analysis, we saw two clear trends. We first discuss the changes with temperature. The data in (Figs 7 & 8) shows that as temperature increases, the rotational correlation times decrease. With use of (Eq. 7.4.1) these correlation times give an estimate of the corresponding rotational hydrodynamic radius. (Fig.17) shows that the calculated hydrodynamic radii had no significant dependence on temperature in the range of dilute concentrations we studied. Noting that (Eq. 7.4.1) has viscosity as a factor, η_W, it is likely that the relationship trend between τ_c values and temperature is mostly attributed to the associated change in solvent viscosity over that temperature range.

We also saw a trend between rotational correlation times and protein concentration. As the protein concentration increased, the rotational correlation times also increased. We hypothesized that increased concentrations of γB-crystallin increased the likelihood that two or more protein molecules were close enough in solution to associate with each other. Once associated, the overall size of the protein will increase, therefore increasing the rotational correlation time, or the time it takes to complete a full rotation in solution. In other words, the τ_c data suggests that increased protein concentrations led to more protein-protein associations. Again, these data were collected at protein concentrations far below *in vivo* concentrations, but we can glean insight into how proteins begin to associate under different conditions.

4.3 Single Vs. Double Exponential Analysis

After determining that the T_1 decay seemed to follow a double exponential decay rather than a single exponential, various mathematical tools were used to test this possibility. First, a

Mathematica script was written to analyze the raw relaxation data for both the T_1 and T_2 experiments. The analysis provided some important numerical results for determining how well an analytical fit matches with a set of data, such as R^2 values, a Durbin-Watson statistic[42], and a residual plot. The analysis was performed assuming both a single and double exponential decay for the T_1 relaxation values at 3 different concentrations and 3 different temperatures.

First, the residual points were plotted to look for patterns indicating autocorrelation. When the residual points were plotted, the single exponential decay yielded residual values 5x greater than the single exponential decay residuals. Additionally, it is apparent from the plot that a sinusoidal pattern is occurring with the residuals for the single exponential decay, while no such pattern is apparent for the double exponential decay. This also suggests the presence of autocorrelation for a single exponential decay, suggesting the double exponential decay fits better.

Then, R^2 values were calculated for all of the relaxations using both a single and double exponential decay. All of the R^2 values were above 0.99, but the double exponential decay fitting yielded values above 0.999 consistently.

The Durbin-Watson statistic was used to test for delay-time autocorrelation in the residual points. Typically, a Durbin-Watson value of 2 indicates no autocorrelation, while values closer to 0 and 4 suggest varying degrees of positive and negative autocorrelation within the residuals. (Table 6) has the Durbin Watson values for the various concentrations and temperatures. Single exponential d values all fell below .57 suggesting a strong positive autocorrelation. One exception to this is the 15°C 8 mg/ml sample, which yielded a d value of 3.18, indicating a strong negative autocorrelation, but the error for this data point is substantially larger than the actual data point, so this data point should not be of too much concern. When a

double exponential decay was fit, d values ranged from 1.8 to 3.1, with a majority of the values falling within .5 of 2, suggesting much less autocorrelation, and a better fit. One thing that needs to be considered regarding the Durbin Watson analysis performed for this analysis is the number of data points being analyzed. For each T_1 analysis, signal strength was analyzed at thirteen different relaxation times, creating a sample size of thirteen for the Durbin Watson analysis. This means that it is likely there is noise that could have an impact on the d value.

Lastly, the Akaike Information Criterion (AIC) was calculated for all available T_1 relaxations. This analytical tool was used for multiple purposes. First, it was used to determine whether a single or double exponential decay fit the data better. At the same time, it was used to assure that we were not overfitting the data by using a double exponential decay rather than a single exponential decay. Finally, the *relative likelihood (i)* was calculated, which provides us with the relative likelihood of one model fitting the data better than the other. In all cases with the exception of the 8 mg/ml sample at 15°C, which favored a single exponential decay, the relative likelihood of the single exponential fitting the data better than the double was 1.45×10^{-5}, and lower. However, the 8 mg/ml sample recorded at 15°C also had the largest amount of error associated with the measurement, so although it is an exception to be noted from the AIC analysis, one should not place too much weight on its importance. The AIC analysis strong indicator that the double exponential decay is a stronger fit then the single exponential decay. Additionally, it provides evidence that a double exponential decay is not an overfit in comparison to the single exponential decay. When used in conjunction with the Durbin-Watson statistical analysis, we are able to build a strong case that the double exponential decay fits the data much better than the single, without forcing an overfit of the data.

Another interesting aspect of the double exponential decay analysis was the second time constant for T_1 NMR. This value was much shorter than previously seen for T_1 relaxation times, especially for biological systems. The origin of this shorter time constant is not yet fully understood, although we hypothesize that it may be the result of a second population of T_1 relaxations occurring on the molecule. The native structure of the γB-crystallin has two individual domains, both of which are of similar molecular weight and structure. The two domains are linked together by a small strand of amino acids. It is possible that the separate relaxation populations we observed were the result of independent movement of these two domains. Another possibility is that the faster relaxation population is actually the result of the C-terminus, or the end of the amino acid chain, rapidly "wiggling" about in solution. Either of these ideas may provide a plausible explanation for the two separate time constants and resulting τ_c times. To corroborate this theory, we used the faster of the two T_1 values (and the experimental T_2 value) to calculate the rotational correlation time using the same τ_c approximation formula implemented by the macro package on the Varian instrument. We further found that this increased rotational correlation time was appropriate for a spherical molecule with a hydrodynamic radius of approximately half that of the normal γB-crystallin. This analysis supports the hypothesis that one of the γB-crystallin domains is exhibiting rotational motions at a rate that is faster than the global protein as a whole.

To further test this hypothesis, the HSQC NMR experimental data will need to be analyzed. This data will provide T_1 relaxation times of specific peaks and allow us to search for residues that are relaxing faster or slower than the rest of the molecule. By determining which part of the protein this different population of residues forms, we can determine if the second

time constant from the double exponential decay is in fact the result of the domains relaxing at separate rates.

When the longer of the time constants from the double exponential analysis was used to calculate the τ_c time using the approximation found in (Eq. 7.4.2), the resulting τ_c times were slower than when a single exponential decay was used, as seen in (Fig. 10). This is to be expected based on how T_1 time constants affect τ_c values. We expect that this longer time constant is representative of the portion of the protein that is tumbling slower than the other portion. Additionally, it is important to note that the single exponential decay time constant appears to be a rough estimation of the weighted averages of the two separate time constants from the double exponential decay.

4.4 Gradient Calibration

All DOSY experiments were performed at the Rochester Institute of Technology on the 500MHz Bruker Avance III 500 spectrometer. It was important to perform a gradient calibration prior to analyzing the DOSY NMR data for our protein. Modern day gradient production methods generally result in reproducible gradients during assembly. However, small deviations in gradient strength may result in errors in diffusion measurements, especially for measurements of molecules that are far from the center of the sample. To calibrate the gradient strengths, solutions with well-known diffusion coefficients can be used. A known diffusion coefficient can be substituted into the Stejskal-Tanner (Eq. 7.4.3), allowing for back-calculation of the actual gradient strength of the probe. Once calculated, the gradient strength is saved into a data file specific to the probe, and automatically applied for all future data analysis. We chose to use a 4.64mM solution of sucrose in D_2O as our calibration standard. The instrument calculated a

diffusion coefficient of 4.05 x 10^{-6} cm^2/s. The published diffusion coefficient for a 4.64mM Glucose in D_2O sample is 4.92 x 10^{-6} cm^2/s. [46] Although these values are not exactly the same, the difference between them (8.7 x 10^{-7} m^2/s) was relatively small. Using the raw data to create a Stejskal-Tanner plot, we found the diffusion coefficient to be 3.77x10^{-6} cm^2/s. Due to time limitations, the gradient calibration was put on hold, and the diffusion measurements for the γB-crystallin are here reported as **pre-calibration values.** As stated earlier, the gradient calibration does not affect how the data are collected, but rather how they are analyzed. This means that the effect of a gradient calibration can be applied post-data collection. Accordingly, future work on this project will require additional work on the gradient calibration using the Stejskal-Tanner equation using solutions with reliable, published diffusion coefficients. Although the gradient was not successfully calibrated, the total difference between the published diffusion coefficient and our value is relatively small (8.66 x 10^{-11}), suggesting that the calibration will only result in a small change to our data. Due to modern day production techniques used when building NMR instruments, gradient calibration is not as important as it once was, and since the magnet and probe this data were collected on are relatively new, it stands that the gradient calibration will not have an overly large impact on our diffusion analysis.

The second issue is a result of the water suppression issue. When we collected the raw data, we noticed that the instrument was eliminating data points from each of the sections

4.5 Translational Diffusion

Translational diffusion measurements were taken over a concentration range between 33 mg/ml and 4.5 mg/ml at 25°C. The signal to noise ratio at the lowest concentration was too low for analysis, so the last data point had to be omitted. After performing the DOSY experiment, the

Bruker software would automatically calculate the resulting diffusion coefficients from the experimental data. The first issue we encountered involving the DOSY experiments occurred during integration. We continually ran into water suppression issues with the γB-crystallin samples. Because of this, when we integrated the 1D spectra, we were forced to integrate upfield and downfield of the water peak separately, rather than integrating over the entirety of the spectra. This resulted in 2 different diffusion coefficients for our samples. However, this is theoretically impossible since translational diffusion is used to measure the longitudinal movement of an entire molecule, rather than just specific regions of the molecule. A literature review of using DOSY to determine diffusion coefficients failed to provide examples of multiple diffusion coefficients for a sample containing a single type of molecule. We believe that reported DOSY diffusion coefficients are the result of averaging the diffusion coefficient over the entirety of the spectra, although further analysis is required.

The second issue we encountered is a result of the water suppression issue. After initially receiving 2 different diffusion coefficients with a difference that was much larger than we expected, we decided to collect the raw data from the instrument. When we began looking at the raw data, we immediately noticed that the instrument was eliminating data points from the analysis. For example, the instrument parameters were originally set up to collected 128 data points from a gradient strength of 2%, up to 98% in equal stepwise increments, but analysis of the raw data from the 33 mg/ml sample showed that only 11 data points were used to calculate the diffusion coefficient of the upfield region, and 33 data points for the downfield region. This means that over half of the data points were eliminated from the instrument's analysis of the diffusion coefficient. Interestingly, each experiment we performed resulted in a different amount of data points eliminated from the analysis. Additionally, the amount of data points used to

calculate the upfield and the downfield region was never the same, which we believe is the cause of the large difference between the resulting diffusion coefficients from the upfield and downfield regions.

The last issue with the DOSY measurements involves error analysis. Unlike the Varian macro package that provided both a τ_c value, as well as the associated error, the Bruker software does not provide any error analysis when it calculates the diffusion coefficient. However, when we calculate the diffusion coefficient using the Stejskal-Tanner plot, the Linest function can be used to determine the error associated with the slope of the line, and therefore the error associated with the diffusion coefficient.

Although we ran into issues that will need to be worked out in before future DOSY analysis can be reliably performed, we were able to collect data to provide insight into the intermolecular behavior of the γB-crystallin. All of the data followed a clear trend, in which the diffusion coefficient decreased as concentration increased. Since the diffusion coefficient is a measurement of distance traveled over time, shorter diffusion coefficients indicate less longitudinal movement of the protein. Our data, therefore, suggest that there are increased intermolecular associations with increased protein concentrations.

5 Conclusion

Brownian dynamics can be used to explain many of the intermolecular interactions between proteins. Here, we studied the effect of various conditions on protein association and/or interactions, with the goal of better understanding how γB-crystallins associate and phase separate *in vivo*. Understanding what causes these phenomena is extremely important for both the treatment and prevention of cataracts. For example, now that a method has been developed to measure and analyze rotational correlation times, γB-crystallins with point mutations could theoretically be purified, and rotational correlation times could be determined to test how specific point mutations affect association rates. This could then be used to develop potential treatment and prevention techniques. Numerous studies have been published regarding Brownian motion of various crystallin proteins.[47] As more of these data are published, a holistic view of the behavior of the crystallin proteins can be developed. Many of these studies focus on the effects of "crowding" and how it impacts the Brownian dynamics. Additionally, this work in conjunction with light scattering measurements of diffusion can be used as a model for the study of the Brownian motion of other proteins outside of the crystallin family.

The τ_c value approximation and derivation are both good indicators of association and/or aggregation rates, although they do not completely describe how the molecules are rotating. An assumption used in our calculations is that the proteins are spherical in shape. As the proteins begin to associate, the shape of the complex does not remain a sphere. To fully consider this aspect of structural change, new τ_c calculations must be derived to consider non-spherical shapes such as cylinders and dumbbell-shaped complexes. Non-spherical shapes will "tumble" in very

different ways than spheres, and the calculations to determine these behaviors are more complex but required to fully understand these motions.

Although a wide range of protein concentrations were studied in this work, our experiments are not representative of *in-vivo* conditions. Even with the development of a new purification method, protein concentrations still fell well below the concentrations found in the human eye. Although the purification process produced pure protein samples, the overall yields were low, likely due to a combination of factors, including protein loss at concentration and filtering steps and an occasional issue with column overloading.

Another important finding of the data that was collected was the discovery of a double exponential decay. As mentioned earlier, generally T_1 NMR is analyzed as a single exponential decay. Using the Durbin-Watson statistic, along with plots displaying the fitted lines, we have developed a strong case in support of a double exponential decay. Additionally, the results of our Akaike Information Criterion analysis support these findings, suggesting that a double exponential decay is a better fit, as well as providing assurance that we are not overfitting the data. We suggest the possibility that spectroscopists are actually seeing the average of two or more relaxation populations when using a single exponential decay. This important revelation could lead to closer attention to the fit of the raw data, allowing for the more detailed analysis of changes in structure and dynamics

Finally, although the gradient calibration provided inconclusive results, it has laid the groundwork for future work. The diffusion measurements collected for our γB-crystallin samples may be adjusted slightly when a better gradient calibration is performed, but the trends will not change. The data clearly suggest that the translational diffusion of the protein is affected by protein concentration. Future work will require more experiments at higher concentrations, as

well as at various temperatures in conjunction with more T_1/T_2 data. Once more NMR data are collected, the long-term goal of this project will be to focus on γB-crystallin with single point mutations of biological interest.

6 References

(1) Andley, U. P. The Lens Epithelium: Focus on the Expression and Function of the Alpha-Crystallin Chaperones. *Int. J. Biochem. Cell Biol.* **2008**, *40* (3), 317–323. https://doi.org/10.1016/j.biocel.2007.10.034.

(2) Delaye, M.; Tardieu, A. Short-Range Order of Crystallin Proteins Accounts for Eye Lens Transparency. *Nature* **1983**, *302* (5907), 415. https://doi.org/10.1038/302415a0.

(3) Common Eye Disorders | Basics | VHI | CDC https://www.cdc.gov/visionhealth/basics/ced/index.html (accessed Feb 14, 2019).

(4) Hejtmancik, J. F. Congenital Cataracts and Their Molecular Genetics. *Semin. Cell Dev. Biol.* **2008**, *19* (2), 134–149. https://doi.org/10.1016/j.semcdb.2007.10.003.

(5) Vision impairment and blindness https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment (accessed Feb 14, 2019).

(6) Chua, J.; Koh, J. Y.; Tan, A. G.; Zhao, W.; Lamoureux, E.; Mitchell, P.; Wang, J. J.; Wong, T. Y.; Cheng, C.-Y. Ancestry, Socioeconomic Status, and Age-Related Cataract in Asians: The Singapore Epidemiology of Eye Diseases Study. *Ophthalmology* **2015**, *122* (11), 2169–2178. https://doi.org/10.1016/j.ophtha.2015.06.052.

(7) Cataracts | National Eye Institute https://nei.nih.gov/eyedata/cataract (accessed Feb 22, 2019).

(8) Wang, J. J.; Mitchell, P.; Simpson, J. M.; Cumming, R. G.; Smith, W. Visual Impairment, Age-Related Cataract, and Mortality. *Arch. Ophthalmol. Chic. Ill 1960* **2001**, *119* (8), 1186–1190.

(9) Bureau, U. C. Older People Projected to Outnumber Children https://www.census.gov/newsroom/press-releases/2018/cb18-41-population-projections.html (accessed Feb 22, 2019).

(10) Wong, T. Y.; Klein, B. E. K.; Klein, R. The Prevalence and 5-Year Incidence of Ocular Trauma11The Authors Have No Proprietary Interest in the Methods or Materials Described in This Paper.: The Beaver Dam Eye Study. *Ophthalmology* **2000**, *107* (12), 2196–2202. https://doi.org/10.1016/S0161-6420(00)00390-0.

(11) Bianco, M.; Vaiano, A. S.; Colella, F.; Coccimiglio, F.; Moscetti, M.; Palmieri, V.; Focosi, F.; Zeppilli, P. Ocular Complications of Boxing. *Br. J. Sports Med.* **2005**, *39* (2), 70–74. https://doi.org/10.1136/bjsm.2003.006643.

(12) Duke-Elder, S. *System of Ophthalmology*; St. Louis, Mosby, 1958.

(13) Ocular Trauma: Acute Evaluation, Cataract, Glaucoma - EyeWiki http://eyewiki.aao.org/Ocular_Trauma%3A_Acute_Evaluation,_Cataract,_Glaucoma (accessed Nov 14, 2017).

(14) McManus, J. J.; Lomakin, A.; Ogun, O.; Pande, A.; Basan, M.; Pande, J.; Benedek, G. B. Altered Phase Diagram Due to a Single Point Mutation in Human ΓD-Crystallin. *Proc. Natl. Acad. Sci.* **2007**, *104* (43), 16856–16861. https://doi.org/10.1073/pnas.0707412104.

(15) Pande, A.; Annunziata, O.; Asherie, N.; Ogun, O.; Benedek, G. B.; Pande, J. Decrease in Protein Solubility and Cataract Formation Caused by the Pro23 to Thr Mutation in Human ΓD-Crystallin,. *Biochemistry* **2005**, *44* (7), 2491–2500. https://doi.org/10.1021/bi0479611.

(16) Dolin, P. J. Ultraviolet Radiation and Cataract: A Review of the Epidemiological Evidence. *Br. J. Ophthalmol.* **1994**, *78* (6), 478–482. https://doi.org/10.1136/bjo.78.6.478.

(17) Hightower, K.; McCready, J. Mechanisms Involved in Cataract Development Following Near-Ultraviolet Radiation of Cultured Lenses. *Curr. Eye Res.* **1992**, *11* (7), 679–689.

(18) Patel, A. S. B - Intraocular Lens Implants: A Scientific Perspective. In *Biomaterials Science (Third Edition)*; Ratner, B. D., Hoffman, A. S., Schoen, F. J., Lemons, J. E., Eds.; Academic Press, 2013; pp 917–930. https://doi.org/10.1016/B978-0-08-087780-8.00078-4.

(19) Phacoemulsification With Intraocular Lens (IOL) Implantation: Background, Indications, Contraindications. **2019**.

(20) MD, R. L. Thoughts on Cataract Surgery: 2015 https://www.reviewofophthalmology.com/article/thoughts-on--cataract-surgery-2015 (accessed Feb 26, 2019).

(21) Cataract Statistics & Resources. *Laser Eye Surgery Hub*, 2017.

(22) Chan, C.-C. Couching for Cataracts in China. *Surv. Ophthalmol.* **2010**, *55* (4), 393–398. https://doi.org/10.1016/j.survophthal.2010.02.001.

(23) ASUQUO, I. M.; BUSUYI, H. M.; UMAR, K. O. The Dangers of Couching in Southwest Nigeria. *Malays. J. Med. Sci. MJMS* **2014**, *21* (5), 60–65.

(24) Ascaso, F. J.; Huerva, V. The History of Cataract Surgery. *Cataract Surg.* **2013**. https://doi.org/10.5772/19243.

(25) Tabin, G.; Chen, M.; Espandar, L. Cataract Surgery for the Developing World. *Curr. Opin. Ophthalmol.* **2008**, *19* (1), 55–59. https://doi.org/10.1097/ICU.0b013e3282f154bd.

(26) Morner, C. T. Untersuchungen Der Proteinsubstanzen in Den Lichtbrechenden Medien Des Auges. *Z Physiol Chem* **1893**, *18*, 61–106.

(27) Wistow, G. The Human Crystallin Gene Families. *Hum. Genomics* **2012**, *6* (1), 26. https://doi.org/10.1186/1479-7364-6-26.

(28) Vérétout, F.; Delaye, M.; Tardieu, A. Molecular Basis of Eye Lens Transparency: Osmotic Pressure and X-Ray Analysis of α-Crystallin Solutions. *J. Mol. Biol.* **1989**, *205* (4), 713–728. https://doi.org/10.1016/0022-2836(89)90316-1.

(29) Horwitz, J.; Bova, M. P.; Ding, L.-L.; Haley, D. A.; Stewart, P. L. Lens α-Crystallin: Function and Structure. *Eye* **1999**, *13* (3b), 403–408. https://doi.org/10.1038/eye.1999.114.

(30) Horwitz, J. Alpha-Crystallin Can Function as a Molecular Chaperone. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (21), 10449–10453.

(31) Bhat, S. P.; Nagineni, C. N. AB Subunit of Lens-Specific Protein α-Crystallin Is Present in Other Ocular and Non-Ocular Tissues. *Biochem. Biophys. Res. Commun.* **1989**, *158* (1), 319–325. https://doi.org/10.1016/S0006-291X(89)80215-3.

(32) Graw, J. Genetics of Crystallins: Cataract and Beyond. *Exp. Eye Res.* **2009**, *88* (2), 173–189. https://doi.org/10.1016/j.exer.2008.10.011.

(33) Vendra, V. P. R.; Khan, I.; Chandani, S.; Muniyandi, A.; Balasubramanian, D. Gamma Crystallins of the Human Eye Lens. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2016**, *1860* (1, Part B), 333–343. https://doi.org/10.1016/j.bbagen.2015.06.007.

(34) Chen, Y.; Zhao, H.; Schuck, P.; Wistow, G. Solution Properties of γ-Crystallins: Compact Structure and Low Frictional Ratio Are Conserved Properties of Diverse γ-Crystallins. *Protein Sci.* **2014**, *23* (1), 76–87. https://doi.org/10.1002/pro.2395.

(35) Kosinski-Collins, M. S.; King, J. In Vitro Unfolding, Refolding, and Polymerization of Human ΓD Crystallin, a Protein Involved in Cataract Formation. *Protein Sci. Publ. Protein Soc.* **2003**, *12* (3), 480–490.

(36) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(37) Chemistry, I. U. of P. and A. IUPAC Gold Book - rotational diffusion https://goldbook.iupac.org/html/R/R05410.html (accessed Feb 28, 2019). https://doi.org/10.1351/goldbook.R05410.

(38) Yao, L.; Vögeli, B.; Ying, J.; Bax, A. NMR Determination of Amide N−H Equilibrium Bond Length from Concerted Dipolar Coupling Measurements. *J. Am. Chem. Soc.* **2008**, *130* (49), 16518–16520. https://doi.org/10.1021/ja805654f.

(39) NMR Relaxation http://chem.ch.huji.ac.il/nmr/techniques/other/t1t2/t1t2.html (accessed Mar 1, 2019).

(40) Petrášek, Z.; Schwille, P. Precise Measurement of Diffusion Coefficients Using Scanning Fluorescence Correlation Spectroscopy. *Biophys. J.* **2008**, *94* (4), 1437–1448. https://doi.org/10.1529/biophysj.107.108811.

(41) Kiraly, P.; Swan, I.; Nilsson, M.; Morris, G. A. Improving Accuracy in DOSY and Diffusion Measurements Using Triaxial Field Gradients. *J. Magn. Reson.* **2016**, *270*, 24–30. https://doi.org/10.1016/j.jmr.2016.06.011.

(42) Durbin, J.; Watson, G. TESTING FOR SERIAL CORRELATION IN LEAST SQUARES REGRESSION. I. *Biometrika* **1950**, *37* (3–4), 409–428.

(43) Akaike, H. A New Look at the Statistical Model Identification. *IEEE Trans. Autom. Control* **1974**, *19* (6), 716–723. https://doi.org/10.1109/TAC.1974.1100705.

(44) Farrow, N. A.; Zhang, O.; Szabo, A.; Torchia, D. A.; Kay, L. E. Spectral Density Function Mapping Using 15N Relaxation Data Exclusively. *J. Biomol. NMR* **1995**, *6* (2), 153–162. https://doi.org/10.1007/BF00211779.

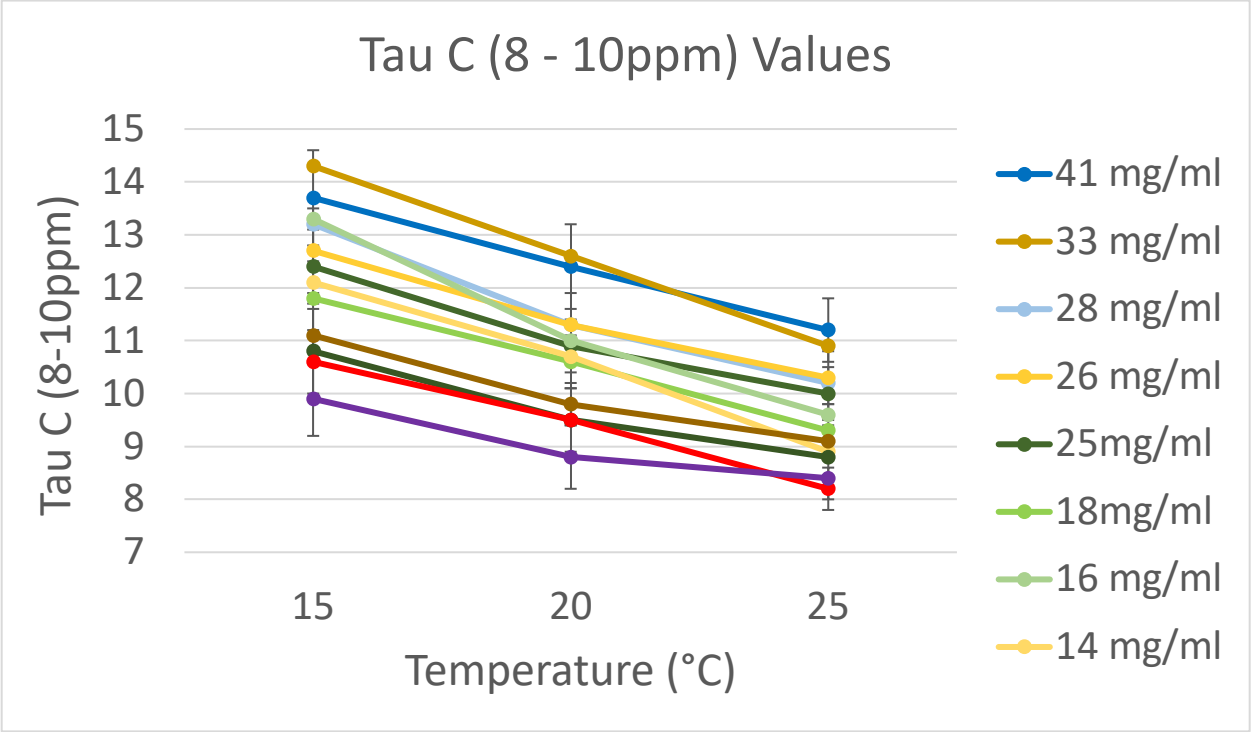
(45) Harris, D. C. *Quantitative Chemical Analysis*; Macmillan, 2010.

(46) Lucas, L. H.; Otto, W. H.; Larive, C. K. The 2D-J-DOSY Experiment: Resolving Diffusion Coefficients in Mixtures. *J. Magn. Reson.* **2002**, *156* (1), 138–145. https://doi.org/10.1006/jmre.2002.2536.

(47) αB subunit of lens-specific protein α-crystallin is present in other ocular and non-ocular tissues - ScienceDirect https://www.sciencedirect.com/science/article/pii/S0006291X89802153 (accessed Feb 21, 2019).

7 Supplemental

7.1 Other τ_c Values recorded over the past 4 years



7.2 Time Constants

7.2.1 T_1 Single Exponential Time Constants

Concentration	Temperature	Time Constant
33mg/ml	25	0.719
33mg/ml	20	0.867
33mg/ml	15	0.986
16mg/ml	25	0.680
16mg/ml	20	0.748
16mg/ml	15	0.900
8mg/ml	25	0.676
8mg/ml	20	0.757
8mg/ml	15	0.924

7.2.2 T_1 Double Exponential Time Constants

Concentration	Temperature	Time Constant 1	Time Constant 2
33mg/ml	25	1.038	0.173
33mg/ml	20	1.210	0.175
33mg/ml	15	1.436	0.166
16mg/ml	25	0.928	0.173
16mg/ml	20	1.058	0.186
16mg/ml	15	1.222	0.163
8mg/ml	25	0.915	0.155
8mg/ml	20	1.003	0.168
8mg/ml	15	0.962	0.030

7.3 Durbin Watson Analysis

7.3.1 Single Exponential Decay

Single Exponential		
Concentration	Temperature	Durbin Watson
33mg/ml	25	0.513
33mg/ml	20	0.513
33mg/ml	15	0.458
16mg/ml	25	0.514
16mg/ml	20	0.503
16mg/ml	15	0.466
8mg/ml	25	0.525
8mg/ml	20	0.567
8mg/ml	15	3.183

7.3.2 Double Exponential Decay

Double Exponential		
Concentration	Temperature	Durbin Watson
33mg/ml	25	1.800
33mg/ml	20	2.553
33mg/ml	15	2.520
16mg/ml	25	3.131
16mg/ml	20	2.289
16mg/ml	15	1.854
8mg/ml	25	1.927
8mg/ml	20	1.949
8mg/ml	15	3.257

7.4 Equations

7.4.1 Hydrodynamic Radius Approximation:

Where:
τ_c=Rotational correlation time
r_H=Effective Hydrodynamic Radius
η_w=viscosity of the solvent
k_B=Boltzmann Constant
T=Temperature

$$\tau_c = \frac{4\pi\eta_w r_H^3}{3k_B T}$$

7.4.2 τ_c Approximation used on Varian Software

$$\tau_c \approx \frac{1}{4\pi\nu_N}\sqrt{6\frac{T_1}{T_2} - 7}$$

Where:
τ_c=Rotational Correlation Time
$\nu_{N=}{}^{15}$N Resonance Frequency (Hz)

7.4.3 Stejskal-Tanner Equation

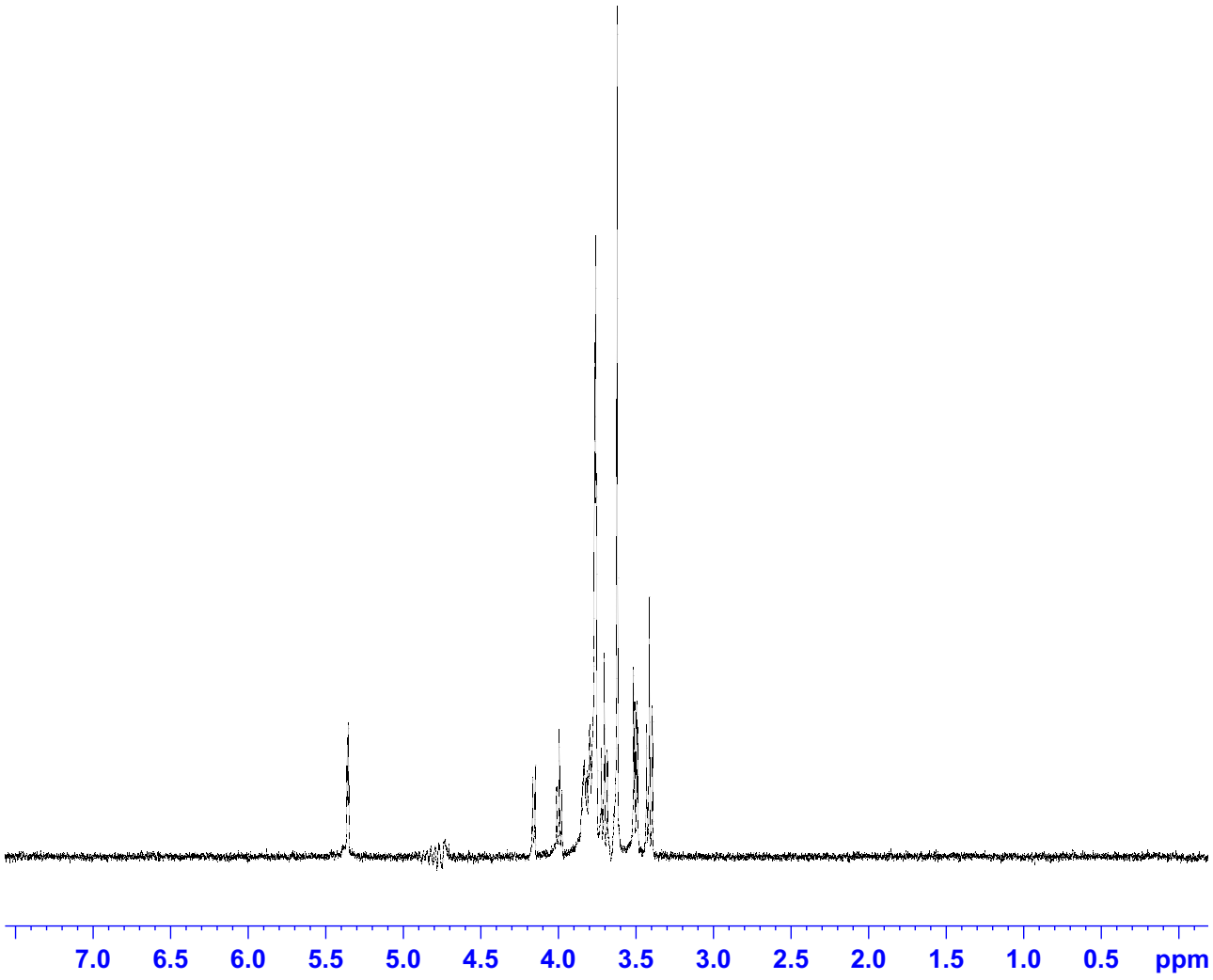
$$I = I_0 e^{-\gamma^2\delta^2 D(\Delta - \frac{\delta}{3})g^2}$$

Where:
γ=gyromagnetic ratio
δ=Gradient length

Δ=diffusion time
G=gradient strength
D=Diffusion coefficient

7.5 Spectra

7.5.1 DOSY Spectrum of Sucrose at 2% Gradient Strength



7.5.2 DOSY Spectrum of 33 mg/ml sample of γB-crystallin at 2% Gradient Strength